

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 20:06:00 ; Search time 3647 Seconds  
(without alignments)  
10916.545 Million cell updates/sec

Title: US-09-943-108a-1

Perfect score: 1368

Sequence: 1 atgcgcattgaaggcttacc.....gtatgaattaccggttttaa 1368

Scoring table: IDENTITY NUC

Gap 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.ov.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1368	100.0	1368	6	E36050	E36050 Signal reco
2	1368	100.0	303750	1	AP003133	AP003133 Staphyloc
3	1368	100.0	347235	1	AP003361	AP003361 Staphyloc
4	1358.4	99.3	295350	1	AP004826	AP004826 Staphyloc
5	790.4	57.8	792	6	E36051	E36051 Signal reco
6	700.8	51.2	319630	6	AX413016	AX413016 Sequence
7	700.8	51.2	349980	6	AX417046	AX417046 Sequence
8	697.6	51.0	2256	6	AX416333	AX416333 Sequence
9	622.8	45.5	2739	1	D14356	D14356 Bacillus su
10	622.8	45.5	208780	1	ESUB0009	Z99112 Bacillus su
11	567	41.4	11023	1	AE006391	AE006391 Lactococc
12	564	41.2	303249	1	AP001515	AP001515 Bacillus
13	522.8	38.2	296750	1	AP003191	AP003191 Clostridi
14	499.2	36.5	10173	1	AE010040	AE010040 Streptoco
15	497.2	36.3	4152	1	U88582	U88582 Streptococ
16	496	36.3	1569	6	AR081886	AR081886 Sequence
17	496	36.3	1569	6	AR195303	AR195303 Sequence
18	496	36.3	1569	6	E35383	E35383 Novel fff.
19	496	36.3	1605	6	AX194269	AX194269 Sequence
20	496	36.3	7577	6	BD003726	BD003726 Polynucle
21	496	36.3	9985	1	AE007428	AE007428 Streptoco
22	496	36.3	10462	1	AE008489	AE008489 Streptoco
23	496	36.3	50946	1	AE014153	AE014153 Streptoco
24	492.8	36.0	1572	6	AX194043	AX194043 Sequence
25	492.8	36.0	77743	2	SPNEU1910	AL449932 Streptoco
26	491.2	35.9	12187	1	AE006560	AE006560 Streptoco
27	486.4	35.6	1136	6	AX433432	AX433432 Sequence
28	479	35.0	11024	1	AE007684	AE007684 Clostridi
29	465.4	34.0	11825	1	AE010644	AE010644 Fusobacte
30	443.4	32.4	10574	1	U32696	U32696 Haemophilus
31	435.4	31.8	11179	1	AE014114	AE014114 Buchnera
32	434.4	31.8	14571	1	AE013104	AE013104 Thermona
33	433.8	31.7	11632	1	AE006158	AE006158 Pasteurel
34	424.2	31.0	11004	1	AE013692	AE013692 Versinia
35	424.2	31.0	220050	1	A7414156	A7414156 Versinia
36	418.2	30.6	293181	1	AP001119	AP001119 Buchnera
37	416.6	30.5	4371	1	BAP239043	AJ239043 Buchnera
38	409.2	29.9	10566	1	AE004142	AE004142 Vibrio ch
39	404	29.5	92407	6	AX067461	AX067461 Sequence
40	404	29.5	343550	1	AP003587	AP003587 Nostoc sp
41	394.2	28.8	2817	1	MYCSRPM54A	MYCSRPM54A
42	393.6	28.0	321250	1	MPULM02	MPULM02
43	377.4	27.6	143308	1	D90913	D90913 Mycoplasma
44	374.4	27.4	11081	1	AE002108	AE002108 Ureaplasma
45	370.6	27.1	4586	1	ECTRMD	X01818 E. coli trm

# ALIGNMENTS

RESULT 1  
E36050  
LOCUS E36050 1368 bp DNA linear PAT 18-JUN-2001  
DEFINITION Signal recognition particle polypeptide and polynucleotide.  
ACCESSION E36050  
VERSION E36050.1 GI:13022452  
KEYWORDS JP 1999235183-A/1.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1368)  
AUTHORS Michael,T.B.  
TITLE Signal recognition particle polypeptide and polynucleotide  
JOURNAL Patent: JP 1999235183-A 1 31-AUG-1999;  
SMITHKLINE BEECHAM CORP

COMMENT	OS	Unidentified	
	PN	JP 1999235183-A/1	
	PD	31-AUG-1999	
	PF	03-SEP-1998 JP 1998289963	
	PR	03-SEP-1997 US 60/057890, 05-MAR-1998 US 09/035382 FI	
	PM	MICHAEL TERENCE BLACK	
	PC	C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00,	
	PC	A61K31/00,	
	PC	A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/70, A61K35/76,	
	PC	A61K39/085,	
	PC	A61K39/395, A61K39/395, A61K48/00, C07K14/31, C07K16/12, C12N1/15,	
	PC	C12N1/19,	
	PC	C12N1/21, C12N5/10, C12N5/10, C12P21/02, C12Q1/68, G01N33/15, PC	
	PC	G01N33/566,	
	PC	C12P21/08, G01N33/577, (C12N15/09, C12R1:445), C12N15/00, C12N5/00,	
	PC	C12N5/00,	
	PC	(C12N15/00, C12R1:445)	
	CC	Strandedness: Single;	
	CC	Topology: Linear;	
	FT	Key	Location/Qualifiers
	FT	source	1. .1368
	FT		/organism='Unidentified'.
FEATURES		Location/Qualifiers	
source		1. .1368	
		/organism="unidentified"	
		/db_xref="taxon:3264"	
BASE COUNT	521 a	186 c	288 g 373 t
ORIGIN			
	Query Match	100.0%;	Score 1368; DB 6; Length 1368;
	Best Local Similarity	100.0%;	Pred. No. 5.3e-220;
	Matches 1368;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	ATGGCATTTGAAGGCTTATCAGACGCTGTCAGACGAGTCGCAAAATTCGGTGTAAG	60
DB	1	ATGGCATTTGAAGGCTTATCAGACGCTGTCAGACGAGTCGCAAAATTCGGTGTAAG	60
QY	61	GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTATTT	120
DB	61	GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTATTT	120
QY	121	GAGCGTCAGGTAACTTTAAAGTGGTAAAGATTTATTAACACAGTATCAGAACCGCA	180
DB	121	GAGCGTCAGGTAACTTTAAAGTGGTAAAGATTTATTAACACAGTATCAGAACCGCA	180
QY	181	TTAGGTTCCGATGTAATCAATATTAAACACAGGCAAGTATTAAATAGTTCAA	240
DB	181	TTAGGTTCCGATGTAATCAATATTAAACACAGGCAAGTATTAAATAGTTCAA	240
QY	241	GATGAATTAACGAAGTTGATGGGTGGAGAAATACATCGATTAATGTCAAAATAACCA	300
DB	241	GATGAATTAACGAAGTTGATGGGTGGAGAAATACATCGATTAATGTCAAAATAACCA	300
QY	301	CCCTACTGTTTATGATGCTGTTTACAGTGCCTGTTAAACACACACACACACACACAC	360
DB	301	CCCTACTGTTTATGATGCTGTTTACAGTGCCTGTTAAACACACACACACACACACAC	360
QY	361	TTAGCATTTATGATCGTAAAAATACACAAAAACCTATGTTAGTTGCAGCAGATATT	420
DB	361	TTAGCATTTATGATCGTAAAAATACACAAAAACCTATGTTAGTTGCAGCAGATATT	420
QY	421	TATCGTCCAGCGGATAAATCAATTAACACAGTAGGGAACAAATGATATTCCTGTA	480
DB	421	TATCGTCCAGCGGATAAATCAATTAACACAGTAGGGAACAAATGATATTCCTGTA	480
QY	481	TACAGTCAGGAGATCAAGTAAACCCACACAAATTTGTAATGATCAATTAACATGCT	540
DB	481	TACAGTCAGGAGATCAAGTAAACCCACACAAATTTGTAATGATCAATTAACATGCT	540
QY	541	AAAGAGACACATTTAGCTTTGATCATGATACAGAGGTCGATACACATCGATGAA	600
DB	541	AAAGAGACACATTTAGCTTTGATCATGATACAGAGGTCGATACACATCGATGAA	600

QY	601	GCATTGATGAACGAATTAAGAAAGTAAAGAAATTCCTAAACCAACGAAATATGTGA	660
DB	601	GCATTGATGAACGAATTAAGAAAGTAAAGAAATTCCTAAACCAACGAAATATGTGA	660
QY	661	GTTGTCGATCAATGAACGGGTCAAGATGCTGCAATGTGCAGAAATCTTTGACGATCAA	720
DB	661	GTTGTCGATCAATGAACGGGTCAAGATGCTGCAATGTGCAGAAATCTTTGACGATCAA	720
QY	721	CTTGATGTCACAGGCTGTTACTTAACTAAATAGATGCTGATACACGTCGTCAGCT	780
DB	721	CTTGATGTCACAGGCTGTTACTTAACTAAATAGATGCTGATACACGTCGTCAGCT	780
QY	781	TTATCTATTTCGTTGGTGACACAAAAACCAATTAATTTGTTGATGAGTGAAGATTA	840
DB	781	TTATCTATTTCGTTGGTGACACAAAAACCAATTAATTTGTTGATGAGTGAAGATTA	840
QY	841	GATGTTTAGAGCTATTCATCTGAACGATGATGATGATGATGATGATGATGATGAT	900
DB	841	GATGTTTAGAGCTATTCATCTGAACGATGATGATGATGATGATGATGATGATGAT	900
QY	901	GTCTTAAGTTTAAATTTGAAAGCGCAACAGATGCGATCAAGAAAAAGCAAAAGATTA	960
DB	901	GTCTTAAGTTTAAATTTGAAAGCGCAACAGATGCGATCAAGAAAAAGCAAAAGATTA	960
QY	961	GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATGATGATGATGATGATGAT	1020
DB	961	GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATGATGATGATGATGATGAT	1020
QY	1021	GTCAAAATCTAGGACCACTGGATGATATTAATGAAAAATGATCCAGGTATGATTAATG	1080
DB	1021	GTCAAAATCTAGGACCACTGGATGATATTAATGAAAAATGATCCAGGTATGATTAATG	1080
QY	1081	AAAGGCTAGATAAGCTTAATATGATGATAAGCAAAATTTGATCATATTAAGCGATTATC	1140
DB	1081	AAAGGCTAGATAAGCTTAATATGATGATAAGCAAAATTTGATCATATTAAGCGATTATC	1140
QY	1141	CATCATGACCGCGCTGAAAGAACATCCAGACACATTCGATGATGATGATGATGATGAT	1200
DB	1141	CAGTCAATGACCGCGCTGAAAGAACATCCAGACACATTCGATGATGATGATGATGAT	1200
QY	1201	CGTATTGCTAAAGGCTGCTGCTTCATTACAGAGTCAATCGTTTGTGATGAACATTT	1260
DB	1201	CGTATTGCTAAAGGCTGCTGCTTCATTACAGAGTCAATCGTTTGTGATGAACATTT	1260
QY	1261	AACGATATGAAGAAATGATCAAAATTCATCTGCTGGCGTAAAGGTAAAAAGGTAAA	1320
DB	1261	AACGATATGAAGAAATGATCAAAATTCATCTGCTGGCGTAAAGGTAAAAAGGTAAA	1320
QY	1321	CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTTAA 1368	
DB	1321	CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTTAA 1368	
RESULT 2			
AP003133			
LOCUS	AP003133	30750 bp	DNA linear BCT 02-JUL-2002
DEFINITION	Staphylococcus aureus subsp. aureus N315 genomic DNA, complete		
genome, section 5/10.			
ACCESSION	AP003133	BA000018	
VERSION	AP003133.2	GI:14349175	
KEYWORDS			
SOURCE	Staphylococcus aureus subsp. aureus N315 (sub_species:aureus N315, strain:N315) DNA.		
ORGANISM	Staphylococcus aureus subsp. aureus N315		
REFERENCE	Bacteria; Firmicutes; Bacillales; Staphylococcus.		
AUTHORS	1		
	Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Iian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Di, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.		

**TITLE** Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*  
**JOURNAL** Lancet 357 (9264), 1225-1240 (2001)  
**MEDLINE** 21311952  
**PUBMED** 11418146  
**REFERENCE** 2 (bases 1 to 303750)  
**AUTHORS** Aoki, K., Oouchi, A., Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K. and Kikuchi, H.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (30-JAN-2001) Akio Oouchi, National Institute of Technology and Evaluation, Biotechnology Center, 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
**COMMENT** (E-mail: oouchi@nitech.go.jp, URL: http://www.bio.nite.go.jp/, Tel: 81-3-3481-8423, Fax: 81-3-3481-8424)  
**FEATURES** On Jun 12, 2001 this sequence version replaced gi:13701012.  
**source** Location/Qualifiers  
1. 303750  
/organism="Staphylococcus aureus subsp. aureus N315"  
/strain="N315"  
/sub\_species="aureus N315"  
/db\_xref="taxon:158879"  
202. .1155  
/gene="SA1056"  
202. .1155  
/gene="SA1056"  
/feature="ORFID:SA1056"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB42308.1"  
/db\_xref="GI:13701013"  
/translations="MKKTIGCLLIMLVVAGSGFGNGKLSKSSKESKQETVKKES  
EEEDPDLXELKMKKIGKADPSLDKLDPLMTEKSTNSKGIQGRKDYKELMGKY  
ELADYRTKDSKSSIKDWAFFKGRKIRKVIETHDDVADVDIWIYDDEGKIKNS  
NTPFYAEIMTKDGLKLVASVEPGSVTHKDDAIDYDYSKLLSQTKLDHDKP  
IVHMDFKSDFGNAHFVLSKVIKEVSNYETASDSLK"  
complement(1318. .1521)  
/gene="SA1057"  
complement(1318. .1521)  
/gene="SA1057"  
/feature="ORFID:SA1057"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAB42309.1"  
/db\_xref="GI:13701014"  
/translations="MQVNVYIILLALFLSGFGIHKFYAKNMQGILHLIFCWGIPH  
1820. .2308  
/gene="SA1058"  
1820. .2308  
/gene="SA1058"  
/feature="ORFID:SA1058"  
hypothetical protein, similar to polypeptide deformylase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB42310.1"  
/db\_xref="GI:13701015"  
/translations="MAIKKLVPSHPILTKAQAQVTKFDDSLKRLQDLEPTVYAOEA  
AGLCAPQNSLOVALIDMEGGLQVNPKKIISQSNWETDLEGSTLPLPYGEVTR  
SKAIVVSGNDVNGKVELFAHEDVARMILHILIQMNGIPTTERADRLITDKEVAYFI  
ND"  
2301. .3236  
/gene="SA1059"  
2301. .3236  
/gene="SA1059"  
/feature="ORFID:SA1059"  
/codon\_start=1  
/transl\_table=11  
/product="methionyl-tRNA formyltransferase"  
/protein\_id="BAB42311.1"  
/db\_xref="GI:13701016"

/translation="MTKILPMGPDPSTTVLEMLIAEHDVIAVVTQDPFVGRKAVMT  
PPVKKVAMKYDLFVQYQELSGSELEQLLOLDVLIYTAARFQGLLPESLIALPKLG  
AVNHASLLPKYRGAPAIHQAIIDGQETGITIMYVKILDAGNIISQOAIKENON  
VGTMDKLSVLGADLLKETLPSIIEGTNESVQDDTQATFASNIRREDERISWKPGR  
OVFNQIRGLSPWPVAYTTMDDTNLKIYDAELVETNKINEPGTIETTKKAILVINDN  
EYAFKDMOLAGKKRMLAANLSGAQNTIVGKKLI"  
3155. .4540  
/gene="SA1060"  
3155. .4540  
/gene="SA1060"  
/feature="ORFID:SA1060"  
hypothetical protein, similar to RNA-binding Sun protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB42312.1"  
/db\_xref="GI:13701017"  
/translations="MSWEKENVSCQLFKWCAKHTSREETYMINVRSIAFDTIQDILN  
EGAYSNLRINEVLSNELNAMDALFETIVGTVKRYTLDQFLKPVVTKIKAWVRQ  
LLWMSIYOYVLDKVPNEAIIINEAVEIAKERGGYHNGVNGILRTMBSDLDPFNEI  
ADPKMALEYSMFKWIIDHWAHYGLETEITLQSTLETSTYVANLTRASLDDII  
EKLDQEGYDVEKDHDPYCLHIGGQPIIHSRSFDDGFVSIOKSSMFVAHIMNVRHD  
HVLDCASAPGGKACHIAEVLMPQGVQYDASDHDKIDILINFNTKKLRUTNIKAFQHD  
TKPYDKYDKILVDAPCGSLGVMRHKPEIKYTSQKHIESLVELQLEILENVKNKYI  
GGEITYSTCTEQLNENVIYTLKNNKNEFEPEFQHPITGELVKTQIMPQDFNSDG  
FRTTKRKDN"  
4543. .5637  
/gene="SA1061"  
4543. .5637  
/gene="SA1061"  
/feature="ORFID:SA1061"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAB42313.1"  
/db\_xref="GI:13701018"  
/translations="MITAEKKKKKFLPNFDKQSIYSLRFDENQNLVLCQGGKFRK  
QIPFELYQKRVDSIDENLNSKDLROLLADNETVTTLTVVYKQESKDGITIKELFLQD  
GYTIEVLMRHVDYGVSVTVQVCRIGCTFCSTGLGKRNLEAGEIYVSQVLTFOKA  
LDATERYVSQIVMIGIGEPEPENNDFRIVNDNSLNGARHITVSTSGIIPRY  
DRADEDIQINFAVSHAAKDEVSRLMPINRAYNVEKLEIAIOYQOEKNRRVTEYG  
LFGVNDQLEHARELAHLIKLNCNVNLPVNHVPERNYVTKANDIFRFEKELARLG  
INATIRREGSGSDAACQLRAKEROVETR"  
5644. .6387  
/gene="SA1062"  
5644. .6387  
/gene="SA1062"  
/feature="ORFID:SA1062"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAB42314.1"  
/db\_xref="GI:13701019"  
/translations="MLEAQPTDTGHRKDNEDAGGIFYNOTQOLLVLCQGGKGRKA  
GEVASFVDELSKFRFAENLIEHOAENWLNKLDINFLYHVAENAEYKGMGT  
CVACALFESVYLVANGDSRAYVINSKQIEQITSDHSVYAHVLTGQITPESAPFEZO  
RNITPKVGTGDKRVSPDLFIRLNFIDYILLNSDGLTDYVKNDEIRALLVKEGTIEDH  
GDQLMQLADNHSKDNVTFFILAALEGDKV"  
6384. .8378  
/gene="SA1063"  
6384. .8378  
/gene="SA1063"  
/feature="ORFID:SA1063"  
/codon\_start=1  
/transl\_table=11  
/product="protein kinase"  
/protein\_id="BAB42315.1"  
/db\_xref="GI:13701020"  
/translations="MICKLINERYKIVDKLGGGMSTVYLAEOTILNKAIVAKAIFIP  
PREKETLAREFERVHNSQLSHQNVIMRDVDEEDDCIYLMEYIEGTFLSEYSEH  
GPLSVDTAINTFNQILGIRKHAHMRIVHRDIPONILIDNSNKLKIFDFGLKALASE  
TSLTQTNHVLGTVQYFSEPAQKEAGTDECTDIYSIYLVLYMLQEPGPNGETVYSIA  
IKHQDSVPNVTTDVKRDIPQSLNSVILRATEKANKRYKTIQEMKDDLSLVHENRA  
NEDVYELDKMKTIAVPLKDEKAKHSEKSNQPKRETTQVPIVNGFAHQOQFQKEG

gene	TVYEPKPKKSTRKIVLLSLIFSLIMIALYSEVAMAFNGKVEETPDVIGKSVKEAEQ	
	IFENKNNLKLKISRSVSDOKYPENELIKTTNTGERTVERGDSVDVVISGEPKVKMPNV	
CDS	IGLKEBALOKLJSLGKLDVITIEKVYINQAPKGIIANQSVTANTEIALHDSNLIKLYES	
	LGIKQVVEDEFHKSRSKAKLKEKFKVESKEEYSDDDIDEGDVISQSPKGSVDG	
gene	STISFYVSKKSDSDSVSTTETTESVDYPTGKNDKSQKVYIKDKNDGSGTEKGSFD	
	ITSQRIIDIPRLIEKKGUASYIVKVDGKTVAEREVSDDI"	
CDS	8606..9481	
	/gene="Sal1064"	
gene	8606..9481	
	/gene="Sal1064"	
CDS	/note="ORFID:Sal1064"	
	/codon_start=1	
gene	/transl_table=11	
	/product="conserved hypothetical protein"	
CDS	/protein_id="BAB42316.1"	
	/db_xref="GI:13701021"	
gene	/translation="MKTGRVKSISGYQVDVNGERFNTKPRGLFRKKKSPVVGDIH	
	EFYQINIEGYTHQFERKNEIKRPVPSNIDTIVIVMSAVEPNFSTQLLDRFLVIAHS	
CDS	YQLNARYLVTKKOKTPLEKQELNELIKYENIGYEFTEFGDDDRKKYIVANPAGLI	
	VLSQSGVSGKSTFLNHYRPELNETNDISKSLNRGKHTTRHVELFERONGYIADTGF	
gene	SALDPDHIDKRETKDYFLELNEYGETCKFRNCHIEKPNVKNVHOLEIGNIAFRDTH	
	YLOLFENISNRVRY"	
CDS	9482..10126	
	/gene="cfxE"	
gene	9482..10126	
	/gene="cfxE"	
CDS	/note="ORFID:Sal1065	
	ribulose-5-phosphate 3-epimerase homolog"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/protein_id="BAB42317.1"	
	/db_xref="GI:13701022"	
gene	/translation="MTKLYPSLLSVDFLDLQHLKLEBAGVDGVHFDVMDGQVPMI	
	SIGPILDAVRKGTLPIDVHLMLENPKVILASFABGKDMLSIHVESTPHIRAIOM	
CDS	IKHLDKAGVYINPGRPISOIEPLIDVILVNTVYVNGFGGQSFIDOCVEKTAGLNA	
	IKMERQNFLEVDGVDNTDAKVCVENGATMLVTSFFFKQEDYKYQTALQG"	
Query Match		100.0%; Score 1368; DB 1; Length 303750;
Best Local Similarity		100.0%; Pred. No. 1.4e-220;
Matches 1366; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	ATGGCAATTTGAAGCCTATACAGACGGCTGCAAGCGAGATGCAAAAAATGCGTGGTAAG 60
Db	26306	ATGGCAATTTGAAGCCTATACAGACGGCTGCAAGCGAGATGCAAAAAATGCGTGGTAAG 26365
Qy	61	GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTATT 120
Db	26366	GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTATT 26425
Qy	121	GAGGCTGACGTAACTTTAAAGTGGTAAAAAGATTATTATAACAGTATCAGAACGCGCA 180
Db	26426	GAGGCTGACGTAACTTTAAAGTGGTAAAAAGATTATTATAACAGTATCAGAACGCGCA 26485
Qy	181	TTAGGTTCGATGTAATCCATCATATACACAGGCGCAACAGTTATATAAATAGTCAA 240
Db	26486	TTAGGTTCGATGTAATCCATCATATACACAGGCGCAACAGTTATATAAATAGTCAA 26545
Qy	241	GATGAATTAACGAAGTTGATGGGTGGAGAAATACATCATGATTAATATGTCAAATAAACA 300
Db	26546	GATGAATTAACGAAGTTGATGGGTGGAGAAATACATCATGATTAATATGTCAAATAAACA 26605
Qy	301	CCFACCTGTTGTTATGATGGTTGGTTTACAGGTGCTGGTAAAAACAACACTGCAGGTAA 360
Db	26606	CCTACTGTTGTTATGATGGTTGGTTTACAGGTGCTGGTAAAAACAACACTGCAGGTAA 26665
Qy	361	TTAGCATTTATGATCGGTAAAAAATACACAAAAACCTATGTTAGTTCAGCAGATATT 420
Db	26666	TTAGCATTTATGATCGGTAAAAAATACACAAAAACCTATGTTAGTTCAGCAGATATT 26725
Qy	421	TATCGTCCAGCAGCATAAATCAATTACACACAGTAGGAGCAACAATTGATATTCCTGTA 480
Db	26726	TATCGTCCAGCAGCATAAATCAATTACACACAGTAGGAGCAACAATTGATATTCCTGTA 26785

Qy	481	TACAGTGAGGAGATCAAGTAAAGCCACACAAAATTGTAACATAATCCATTAAAAACATGCT 540
Db	26786	TACAGTGAAGGAGATCAAGTAAAGCCACACAAAATTGTAACATAATCCATTAAAAACATGCT 26845
Qy	541	AAAGAAGAACATTTAGACTTTGTAATCATTATACAGCAGTCGATTACACATCGATGAA 600
Db	26846	AAAGAAGAACATTTAGACTTTGTAATCATTATACAGCAGTCGATTACACATCGATGAA 26905
Qy	601	GCATTGATGAACGAATTAAGAAGTAAAGAAATTTGTAACCAACGAATATATGTTA 660
Db	26906	GCATTGATGAACGAATTAAGAAGTAAAGAAATTTGTAACCAACGAATATATGTTA 26965
Qy	661	GTGTGCGATTCAATGACGGGTCAAGATGCTCAATGTTGCAGATCTTTTGACGATCAA 720
Db	26966	GTGTGCGATTCAATGACGGGTCAAGATGCTCAATGTTGCAGATCTTTTGACGATCAA 27025
Qy	721	CTTGATGTCACAGGTGTTTACCTTAACATAATTAGATGGTGATACACGTGGTGGTCAGCT 780
Db	27026	CTTGATGTCACAGGTGTTTACCTTAACATAATTAGATGGTGATACACGTGGTGGTCAGCT 27085
Qy	781	TTATCTATTTCGTTGCGTGACACAAAAACCAATTAATTTGTTGGTATGAGTGAAAAGTTA 840
Db	27086	TTATCTATTTCGTTGCGTGACACAAAAACCAATTAATTTGTTGGTATGAGTGAAAAGTTA 27145
Qy	841	GATGGTTTAGAGCTATTCATCCATCGTGAACGATATGGCATCACGTCATTTTAGTATGGTGAT 900
Db	27146	GATGGTTTAGAGCTATTCATCCATCGTGAACGATATGGCATCACGTCATTTTAGTATGGTGAT 27205
Qy	901	GTGTTAAGTTTAATTGAAAAAGCGCAACAAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960
Db	27206	GTGTTAAGTTTAATTGAAAAAGCGCAACAAGATGTGGATCAAGAAAAAGCAAAAGATTTA 27265
Qy	961	GAGAAAAAGATCCGTGAGTCATCGTTTACTTTTAGATGATTTTGTAGAACACCTTGATCAG 1020
Db	27266	GAGAAAAAGATCCGTGAGTCATCGTTTACTTTTAGATGATTTTGTAGAACACCTTGATCAG 27325
Qy	1021	GTGAAAAATCTAGGACCACTGGATGATATTATGAAAAATGATTCAGGATGATGAATAAATG 1080
Db	27326	GTGAAAAATCTAGGACCACTGGATGATATTATGAAAAATGATTCAGGATGATGAATAAATG 27385
Qy	1081	AAAGGCTAGATAAGCTTAAATGATGTAAGGAAAGCAAAATGATCATATTAAAGCGATTATC 1140
Db	27386	AAAGGCTAGATAAGCTTAAATGATGTAAGGAAAGCAAAATGATCATATTAAAGCGATTATC 27445
Qy	1141	CAGTCAATCAGCCCGGCTGAAAGAAACAATCCAGACACATTGAATGTATCCGCTAAATAAG 1200
Db	27446	CAGTCAATCAGCCCGGCTGAAAGAAACAATCCAGACACATTGAATGTATCCGCTAAATAAG 27505
Qy	1201	CGTATTGCTAAAGGCTCTGGTCGTTTACAAAGAGTCAATCGTTTGTATGAAACAATTT 1260
Db	27506	CGTATTGCTAAAGGCTCTGGTCGTTTACAAAGAGTCAATCGTTTGTATGAAACAATTT 27565
Qy	1261	AACGATATCAAGAAATGAGCAACAATTCACCTGTCGGGTAAAGGTAAAGGTAAAGGTAA 1320
Db	27566	AACGATATCAAGAAATGAGCAACAATTCACCTGTCGGGTAAAGGTAAAGGTAAAGGTAA 27625
Qy	1321	CGCAATCAAAATGCAAAATATGTTAAAAGGTATGAATTTACCGTTTTAA 1368
Db	27626	CGCAATCAAAATGCAAAATATGTTAAAAGGTATGAATTTACCGTTTTAA 27673

RESULT 3  
AP003361  
LOCUS AP003361 347235 bp DNA linear BCT 07-FEB-2002  
DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete  
sequence, section 4/9.  
ACCESSION AP003361 BA000017  
VERSION AP003361.2 GI:14246761  
KEYWORDS  
SOURCE Staphylococcus aureus subsp. aureus Mu50 (sub\_species:aureus Mu50,  
strain:Mu50) DNA.  
ORGANISM Staphylococcus aureus subsp. aureus Mu50  
Bacteria; Firmicutes; Bacillales; Staphylococcus.



REFERENCE  
AUTHORS

1 Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I., Cui L., Ouchi A., Aoki K., Nagai Y., Lian J., Ito T., Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani Y., Takahashi N.K., Sawano T., Inoue R., Kaito C., Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M., Ogasawara N., Hayashi H. and Hiramatsu K. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*

## TITLE

Lancet 357 (9264), 1225-1240 (2001)

## JOURNAL

## MEDLINE

21311952

## REFERENCE

2 (bases 1 to 347235)

## AUTHORS

## TITLE

Direct Submission  
Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College of Medical Technology and Nursing, Department of Medical Technology, 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan  
(E-mail: tohtasakura.cc.tsukuba.ac.jp, Tel: 81-298-53-3434, Fax: 81-298-53-3454)

## COMMENT

## FEATURES

## SOURCE

1. .347235  
/organism="Staphylococcus aureus subsp. aureus Mu50"

/strain="Mu50"

/sub\_species="aureus Mu50"

/db\_xref="taxon:158878"

35. .1021

/gene="SAV0992"

/gene="SAV0992"

/codon\_start=1

/transl\_table=11

/product="oligopeptide ABC transporter ATP-binding protein homologue"

/protein\_id="BAB57154.1"

/db\_xref="GI:14246762"

/translation="MNNVLEVKDLERSLKINNEWLATVENISPELSKGEVLGVGES CGKSLIKLPEIKSLSSGEVIFDGKRLDTLNEKQLDIRGNDIAMIFQPPM TALNVEITKNQVIESIKSHKSKKANLAKDLKLVGTARQDEILNSPHOLSGG MRQWMTAMAISCPKLTADPTALDVTIQAILDLKELQETQAMIMTHDLS VVAFQCKVLVMYAGQIVFEGGKEILHNPYTKOKLSTIPKLKQKRLFTIBGI VPSTQAPHNKCRCANCKKLDICNNQSPKMHVCEIVIRCHLVKNEYKEI"

1024. .2004

/gene="appf"

1024. .2004

/gene="appf"

/note="SAV0993"

/codon\_start=1

/transl\_table=11

/product="oligopeptide transport system ATP-binding

protein Appf homologue"

/protein\_id="BAB57155.1"

/db\_xref="GI:14246763"

/translation="MENILEVNQIKVYKIKTGLOKTOYKAVDVSFSIKKGQTFG LVGSGGKSTLGKVTIRLEADTSIIIVNGEDITRLOGKILKRSQOQIMQDPVA SLNPMQWGDIISEPIYAKKLPEIKKLYLLKCVGLSEDAVYKXAFHFGGQKQ RVGIARALRPLSVIADPEVSALDVSQCVNLKLDQKLFSLYFIADHLSVVK HISVIGWTLGHVETIASDKEIYENFAHPTKALISSIPOIDKNNRRIILKGLPS PSNPQCPPTTCPIAKDKCKENIPQKIDGNEHQVACFYVDYKVDLNA"

1997. .2959

/gene="oppB"

1997. .2959

/gene="oppB"

/note="SAV0994"

/codon\_start=1

/transl\_table=11

/product="probable oligopeptide transport system permease protein"

/protein\_id="BAB57156.1"

/db\_xref="GI:14246764"

/translation="MLKLIKRLGLMPLLLILISIVFSLAIHQDPFSLQNGKIK QEAINAQREKLGNDLIPHQYIRWVNHVTHGDLGESIKYKRPVIDVIERIPNTILIG

## gene

AMSLITYITISFALGITSGRYSLSLTDYTVQIFNYMLAIPSFIAQVFAIFISPELQ  
WFPGQSVINDLKEGTEYTMKSYHTFLPALTGLLSTAGYIQYLRNDIENSKDY  
VLFARSKSNKYNKHIKRLNSLIPLITFLGADIVSLTGAVTIFSYNGICKLF  
LESVIGQDYLPMALITLFFSFLGLNLSIDITYGFIDPRISN"

## CDS

2971. .3852

/gene="SAV0995"

2971. .3852

/gene="SAV0995"

/note="similar to oligopeptide transport system permease

protein OppC"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAB57157.1"

/db\_xref="GI:14246765"

/translation="WQNKSKPFIKIALSRFHINKTAMLSIIFLITITIVSIITAPIAP  
FPYNODLNLKGEWTAQNLGDSGRDNFSELYAGRLSLSIGTSIGMLLIGIT  
VGVISGFGIVDILMRITHEFVLMFPFLIFALVJNALGDKINKPYGSAIILVLVII  
VLSGGIARLVNGKVLQEKENEYFLAAKSIGTPTTKIILKHLNLSLVVIVQATLLE  
AMIVVSGLSFLGFGISKAIPSGNMLSDAQEGDVISGKFWIWPMPAINITILSI  
NFVGEGLKDAFNPRGR"

complement(3894. .4883)

/gene="trpS"

complement(3894. .4883)

/gene="trpS"

/note="SAV0996"

/codon\_start=1

/transl\_table=11

/product="tryptophanyl-tRNA synthetase"

/protein\_id="BAB57158.1"

/db\_xref="GI:14246766"

/translation="METLFGIQSPGIPIGNVIGALKQFVDVQNDYDCVCFIVDOHA  
ITMPDRLKLRQRTQAAIYASGIDDPKATLFIQSEVPAHVQAGMLTTIASVGEEL  
RNSQDKRAQVAGEIPAGLITPPLMAADIVLTNTIVPGDDQKQHIETLRNLVD  
RFRSYNDVYVKEIRMPKVGGRVMSLQDPTKMSKSDNAKNTISLLDEPNVAAKI  
KSAVTDSDGIIKFRDNKPGTTLISITVAGLTDMPIDKIEAKYEGEGYKFGKGLAEI  
VKAFVFEQKYESPYNSDKLDDLQDQGRKAHKVSEKTVKKMEKAGLGRKR"

5177. .5572

/gene="SAV0997"

5177. .5572

/gene="SAV0997"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="BAB57159.1"

/db\_xref="GI:14246767"

/translation="MVTLFTSPSTCRKAKAWLQEHDPYTERNIFSEHLTIDEIKQ  
ILKMTEDGTDEIISTRSTYQKLNVDISLPLQDLYSIQDNPGLLRRLTILDNKRLQ  
VGYNEDEIRRFPRKRVTFQLQEAQRMD"

5943. .6662

/gene="SAV0998"

5943. .6662

/gene="SAV0998"

/note="similar to negative regulator of genetic competence

MeoA"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAB57160.1"

/db\_xref="GI:14246768"

/translation="MRIERVDVDTVKLITYSDIEARGFSREDLWTKRKEGFEFWSM  
MDRIEEDFVVEGELWIOHAFKGVETVTSKNEDMMMSDDDATDOFQOQVHLS  
LAOTLEGDQLELPEQRTKEKQSKRSKARNTRTIIIVKFNDELVINYVHLS  
NPITTEFDLLYWDGTYTYAYAFHDSHVQDEVINDYSQLLERFAYPTDRTVYLNDYA  
KIIMSHNVTAQVRRYFPETTE"

6783. .7769

/gene="SAV0999"

6783. .7769

/gene="SAV0999"

/note="similar to transcription factor"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

gene  
/protein\_id="BAB57161.1"  
/db\_xref="GI:14246769"  
/translation="MLVALNEKERVLIATLRTQVFCVCGKQVILKRGKLVISHF  
AKHLAEOKCFNNETIKHKSKLILLAQMLQQGCKVELEPFLEIKIQIDILNNKYV  
IELQSPYKQILQRTREGKMKGYKSWLLNDVDYCHNKYFNHFSLFNIPTRKL  
HFNLEKQIMFQIQIYLGHHYVAERKNKAKIIELENEAPCDYHAYVYKLSFALNOY  
IKYKRVNLEPTLSAMQLOLTDOEVYVNGYIFPEQYIENHPIEWQLQVLDWLK  
NGSKLVNDNLNFKLKFIVALESKTAIEKLINNLINICSDRGNDVQILE"  
7817..9625  
/gene="SAV1000"  
7817..9625  
/codon\_start=1  
/gene="SAV1000"  
/transl\_table=11  
/product="thimet oligopeptidase homologue"  
/protein\_id="BAB57162.1"  
/db\_xref="GI:14246770"  
/translation="MSQQLSREQEERKPEYTWLITFIKDDERFAAEKFEVENELGK  
EQDFKGHSASFLYNALAELEDTIKLEKYVYVYHLKQDDTNDKRTIGMESRAHQ  
LITKFSWSPLVEILQIDEDIKQSFVNSDKLQKPAFDLKLINERPHILDATKX  
LTERQDALSTPSNVYGMFSNADLVFEDAIDKGNAPFLTQGTFIKLESDDKRLRES  
APRNYKAYGAHNNLTGAILAGEVKNVFNATNYKTAREKALSNHHPENVYDNLV  
KYVXYLLHRYTELKELLGLLDLKMVDLITPLDKIFEMPIEERKEWMLKALEP  
MGEETLVNKKGLNWRVDYENKRRGSGYSSGAHLNFPILLNSNTISDLYTLVE  
EFGHSASFYFRKFPQPSNSDYTIFFVAEYSTCNEALLSDYMDKHLDDKRLILLNQE  
LEREFATFRQTFAPAEPEHKIHAIEAGEPLTPTRMNEEYAKLNKLYFGDSVETEDI  
SKEWRSIPHYMYVYVYATGYSAQSLSHOILTEGKPAVDRYINEFLKGSNNYPI  
ELKNAGVDMTPPEPTEQACEVPEQKLNFEKLMA"  
complement(10085..10639)  
/gene="yjbM"  
complement(10085..10639)  
/gene="yjbM"  
/note="SAV1001"  
/codon\_start=1  
/transl\_table=11  
/product="GTP pyrophosphokinase"  
/protein\_id="BAB57163.1"  
/db\_xref="GI:14246771"  
/translation="MKOLSVKVVSKTEKIYTCYNELIPKDIILITESMICDIONAG  
IDLEVPKDLQKSLPESLKIDLHTAREMEIQAPLSVFPESDVHEEGKVBGLTPYH  
IYIIEINLMGKPEKLNPKPLETYIQQQQLVITMEELLTYEWPEKLLNKLKLAIQ  
KIKELKYPDGDGKSKMPKRSK"  
complement(10575..10901)  
/gene="SAV1002"

Query Match 100.08; Score 1368; DB 1; Length 347235;  
Best Local Similarity 100.08; Pred. No. 1.3e-220;  
Matches 1368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGCATTGAAGGCTTATCAGACGCTGCAAGCGAGCATGCAAAAATGCGTGGTAAG 60  
DB 256730 ATGGCATTGAAGGCTTATCAGACGCTGCAAGCGAGCATGCAAAAATGCGTGGTAAG 256789  
QY 61 GGTAACTTACTGAAGCTGATATAAGATAATATGATGCGTGAAGTAAAGATTAGCGTTATTT 120  
DB 256790 GGTAACTTACTGAAGCTGATATAAGATAATATGATGCGTGAAGTAAAGATTAGCGTTATTT 256849  
QY 121 GAGCGTCAGCTAACTTAAAGTGTGTTAAAGAAATTTATTAACACAGTATCAGACGGCA 180  
DB 256850 GAGCGTCAGCTAACTTAAAGTGTGTTAAAGAAATTTATTAACACAGTATCAGACGGCA 256909  
QY 181 TTAGGTTCCGATGTAATGCAATCATTAACACAGGCGCAACAGTATTATAAATAGTTCAA 240  
DB 256910 TTAGGTTCCGATGTAATGCAATCATTAACACAGGCGCAACAGTATTATAAATAGTTCAA 256969  
QY 241 GATGAATTAACGAAGTGTGTTGGTGGAGAAATATACATCGATTAAATATGTCAAATAACCA 300  
DB 256970 GATGAATTAACGAAGTGTGTTGGTGGAGAAATATACATCGATTAAATATGTCAAATAACCA 257029  
QY 301 CCTACTGTTGTTATGATGTTGGTTTACAGGTGCTGGTAAACACACACTGCAGGTAAA 360  
DB 257030 CCTACTGTTGTTATGATGTTGGTTTACAGGTGCTGGTAAACACACACTGCAGGTAAA 257089

QY 361 TTAGCATTATTGATGCGTAAAAAATACAAAAAACCCTATGTTAGTTCGACGAGATATT 420  
DB 257090 TTAGCATTATTGATGCGTAAAAAATACAAAAAACCCTATGTTAGTTCGACGAGATATT 257149  
QY 421 TATCGTCCAGCAGCGATAAATCAATTTACAAACAGTAGGGAACAAATGATATTCCTGTA 480  
DB 257150 TATCGTCCAGCAGCGATAAATCAATTTACAAACAGTAGGGAACAAATGATATTCCTGTA 257209  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAACAAATTTACTACTAATGATTAACATGCT 540  
DB 257210 TACAGTGAAGGAGATCAAGTAAAGCCACAAACAAATTTACTACTAATGATTAACATGCT 257269  
QY 541 AAAGAAGAACATTTAGACCTTTGTAATCATTCATACACAGGTCGATTTACATCGATGAA 600  
DB 257270 AAAGAAGAACATTTAGACCTTTGTAATCATTCATACACAGGTCGATTTACATCGATGAA 257329  
QY 601 GCATTGATGAACGAATTAAGAAGTAAAGAAATGCTTAACCAACAAAGAAATATATGTTA 660  
DB 257330 GCATTGATGAACGAATTAAGAAGTAAAGAAATGCTTAACCAACAAAGAAATATATGTTA 257389  
QY 661 GTTGTCGATTCAATGACGGGTCAAGATGCTGCTCAATGTTTCAGAAATCTTTTGAAGATCAA 720  
DB 257390 GTTGTCGATTCAATGACGGGTCAAGATGCTGCTCAATGTTTCAGAAATCTTTTGAAGATCAA 257449  
QY 721 CTTGATGTCACAGGTGTTACCTTAATTAATTAATTAATGTTGGTATGAGTGAAGAAATTA 840  
DB 257450 CTTGATGTCACAGGTGTTACCTTAATTAATTAATGTTGGTATGAGTGAAGAAATTA 257509  
QY 781 TTATCTATTCTGTTCCGTGACACAAAAACCAATTAATTAATGTTGGTATGAGTGAAGAAATTA 840  
DB 257510 TTATCTATTCTGTTCCGTGACACAAAAACCAATTAATTAATGTTGGTATGAGTGAAGAAATTA 257569  
QY 841 GATGGTTTAGAGCTATTTCATCTCTGAACGATGATGATGATGATGATGATGATGATGATGAT 900  
DB 257570 GATGGTTTAGAGCTATTTCATCTCTGAACGATGATGATGATGATGATGATGATGATGATGAT 257629  
QY 901 GTGTAAAGTTAAATGAAAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960  
DB 257630 GTGTAAAGTTAAATGAAAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 257689  
QY 961 GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTGTAGAACAACTTGTATCAG 1020  
DB 257690 GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTGTAGAACAACTTGTATCAG 257749  
QY 1021 GTGAAAAATCTAGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
DB 257750 GTGAAAAATCTAGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 257809  
QY 1081 AAAGGGCTAGATAGCTTAAATGATGAAAAAGCAAAATGATCATATTAAGCGATTATC 1140  
DB 257810 AAAGGGCTAGATAGCTTAAATGATGAAAAAGCAAAATGATCATATTAAGCGATTATC 257869  
QY 1141 CAGTCAATGACGCGGCTGAAAGAAACAAATCCAGACATGATGATGATGATGATGATGATGATGATGAT 1200  
DB 257870 CAGTCAATGACGCGGCTGAAAGAAACAAATCCAGACATGATGATGATGATGATGATGATGATGATGAT 257929  
QY 1201 CGTATTCCTAAAGGGCTGCGTTCATTACAAAGCAAGTCAATCGTTTGTATGATGATGATGATGAT 1260  
DB 257930 CGTATTCCTAAAGGGCTGCGTTCATTACAAAGCAAGTCAATCGTTTGTATGATGATGATGATGAT 257989  
QY 1261 AACCATATGAAGAAATGATGAAACAAATTCATCTGTCGGGTAAAGGTAAAAAGGTAAA 1320  
DB 257990 AACCATATGAAGAAATGATGAAACAAATTCATCTGTCGGGTAAAGGTAAAAAGGTAAA 258049  
QY 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1368  
DB 258050 CGCAATCAATGCAAAATATGTTAAAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 258097

RESULT 4  
AP004826 295350 bp DNA linear BCT 02-JUL-2002  
LOCUS AP004826  
DEFINITION Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,

```

strain:MW2, section 5/10.
ACCESSION AP004826 BA000033
VERSION AP004826.1 GI:21204263
KEYWORDS
SOURCE Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.
ORGANISM Staphylococcus aureus subsp. aureus MW2
REFERENCE Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,
Nagai,Y., Iwana,N., Asano,K., Naimi,T., Kuroda,H., Cui,L.,
Yamamoto,K. and Hiramatsu,K.
TITLE Genome and virulence determinants of high virulence
community-acquired MRSA
JOURNAL Lancet 359 (9320), 1819-1827 (2002)
MEDLINE 22040717
PUBMED 12044378
REFERENCE 1
AUTHORS 2 (bases 1 to 295350)
Nagai,Y., Oguchi,A., Nagai,Y., Asano,K., Iwana,N., Baba,T.,
Kuroda,M., Hiramatsu,K. and Kikuchi,H.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2002) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
FEATURES
Source location/Qualifiers
1. .295350
/organism="Staphylococcus aureus subsp. aureus MW2"
/strain="MW2"
/db_xref="taxon:196620"
301..1254
/gene="MW1096"
301..1254
/gene="MW1096"
/notes="ORFID:MW1096"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB94961.1"
/db_xref="GI:21204284"
/translation="MKKTGCLLLMLLVAGSCFSGGNHLSKSKSESKQPTVKES
EEKDPDEIEEIKRMKGIDKSLDLPMTKSFNSKGIQGWKDYKELMGKV
ELADYFTKDSKSGSIDVAFKGGKIKRKVIETHDDVDVYVDPDGKKGINS
NTVYFAETMKYDKGLVYASVPGSVYHKDDAIVKDDYDKLKLSTLKHDPKP
VPSVAQIKSCGVELTSVSPMTGSKDDEVLPALAYTFESPKNYEDKSNPDKVLN
LVHMDPLARSSDFGNHFFVLSKFIKEISNTERASDDSLK"
/gene="MW1097"
complement(1417..1695)
/gene="MW1097"
/notes="ORFID:MW1097"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94962.1"
/db_xref="GI:21204285"
/translation="MFELKVGIIKREANVIISNSTGGKEMOVNKVIYILLALFSGFG
IHKFYAGKNMGILLHFCWTGTPHILAIISAVITVFPKPEQGNVTL"
1919..2407
/gene="MW1098"
1919..2407
/gene="MW1098"
/notes="ORFID:MW1098"
hypothetical protein, similar to polypeptide deformylase"
/codon_start=1
/transl_table=11
/protein_id="BAB94963.1"
/db_xref="GI:21204286"
/translation="MAIKLVPASHPLTKQAQVIKFDSDSLKRLQDLEDNWAQEA
AGLCAPQINQLOVALIDMEMEGLQVNPKLIISOSNETLTDLGGSTLDPDYGVETR
SKMIVVESDVNGNKKVELTAHEDVAMHILHIDQNGIPFTERADRLIDKEVEYFI
ND"
2400..3335
/gene

```

---

```

/gene="MW1099"
2400..3335
/gene="MW1099"
/notes="ORFID:MW1099"
/codon_start=1
/transl_table=11
/product="methionyl-tRNA formyltransferase"
/protein_id="BAB94964.1"
/db_xref="GI:21204267"
/translation="MTKTIEMGTDFSTVLEMLIAEHVDIAVTQDPRVGRKRYMT
PPPKYKAMKIDLPVYQPEKISGSEBEQLQLDVLITVAFGQLLPESLLALPKLG
AINVASLLPKYRGGAPIHQAIIIDEGQETGITIMYVKKLDAAGNIISQAALPEEND
VGTMDKLKSLGDAKLTLPSTIIEGNESVPQDDTQATFASNIRREDRINNNKPGR
QVFNQIRGLSPFVAYTTMDDTNLKIYDAELVETKNIEPGTIIETTKAIIIVTNDN
EAVAIDKMLAGKKGKMLAANTLSGAQNTLVGKKLI"
3332..4639
/gene="MW1100"
3332..4639
/gene="MW1100"
/notes="ORFID:MW1100"
hypothetical protein, similar to RNA-binding Sun protein"
/codon_start=1
/transl_table=11
/protein_id="BAB94965.1"
/db_xref="GI:21204268"
/translation="MIENVRSIAFTTQDILNEGAYSNLRINEVLSSENLAMDKALF
TEIYGVTKRFTLDFLFPFVKTKIRAWRQLLWMSIYQYVYLDKVPNHAIINEAVE
IAKRGGYHNGYVNGILFTMRSDLPFNEIADPKRMAIEYSMPKWIIDWATHYGP
LEKTEFTILQSFLETTSTTVRANLTRASLDODIIEKLQDEGVDEKDDHLPYCLHIGGP
IIHSRSFKDFGVSIQDKSSMFVAHIMNVDRHDVLDACSPAGSKACHIAVLMPEGOV
DASDIDHKIDLLINENIKRLTNKAFQHDATKPYDKTYDKILDADAPCSGLGVMRHK
PEIKYTSKOHIESIVLELOILELNVKNVKNIGELIYSTCTIEQLENEVITYFLAN
NKNFEFPFQHPITGELVKTLQIMQDPNSDGFITKIRKKON"
4642..5736
/gene="MW1101"
4642..5736
/gene="MW1101"
/notes="ORFID:MW1101"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94966.1"
/db_xref="GI:21204269"
/translation="MITAEKKKKKLPNPKOSIYSLRDEMQWLVQEQOKERAK
QIFELKQKVDSDIDENLTKDQLLKQNFVTLTVVKQESKDSITIEFLYELQD
GIIETVLMKHDGNSVCVTIQVGRIGCFPCASIIQGLKRNLEAGEISVSLIVQKA
LDATFERSQVILVINGEPEFFNYDEMDFLRIVNDDNSLNGARHITVSTSGIIPRIY
LFADEDIQINFAYSLHAADKDEYRSLAMPINRAYNVEKLIEATQYQETKRRVTFEYG
LFGVNDQLEHARELAHLIKGLNCHVNLIPYVHPERNYVKTAKNDIKFKEKELRLG
INATIRREGSDIDAACQQLRAKERQVETR"
5743..6486
/gene="MW1102"
5743..6486
/gene="MW1102"
/notes="ORFID:MW1102"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94967.1"
/db_xref="GI:21204270"
/translation="MLEAQFTDQGHDKNEDAGGIYFNRTNQQLVLCDSMGGHKA
GEVASKFVTDELKAREAEHLIEHQENLNRNKNIDINFQLYHYAQENAEYKMGMT
CVCAIVPEKSVIANQDSRAVYINRSQIQITSDHSFVNLHVLVTGQITPEAPTPO
RNIITKVMGDKVSPDLFIKRLNFDYLLNLSDLGTDVYKDNKIRLLVKEGTIEDH
GQDLMLALDNHSDKNVTFLAAIEGKV"
6483..8477
/gene="MW1103"
6483..8477
/gene="MW1103"
/notes="ORFID:MW1103"
hypothetical protein, similar to protein kinase"
/codon_start=1
/transl_table=11

```

/protein\_id="BAB94968.1"  
/db\_xref="GI:21204271"  
/translation="MIGKIINERYKIVDKLGGGMSVYLAEDTILINIKVAIKAFIP  
PREKEETIKFEREVENSQLESHQIVSMIDVEDDDCYLVMEYIEGPTLSYIESH  
GSLVDFAINFTQOILDKKHAHDMRIVHRDIPQNLIDSNKILKIFDGIKALSE  
TSLQTHNLVLTQVYFSPQAKGEATDECDTIVSIGIVLYEMLVGEPFNGETAUSTA  
LKHQDSVPNVTVDDVKDIPQSLSNVILRAPEKDKANVKTIOEMKDDLSVYHENEA  
NEDYELDKMKTIAVDIAKKEDLAKHLSHKSNOQKRETTQVPIVNGPAHQOQFOKPG  
TVYEPKSKTKRKIVLLSLIFSLMIALVSVAMAFNGKYEETPDVIGKSVKEAEG  
IFNNKLNKIKRSYSDKYPENELIKTPNTGERVSGSDVYVISKPEKVKMPNV  
IGLPEKALQKLSLGLKQVTEKYVNNQAPKGYIANOSVTANTEIAIHDNINIKLYES  
LGIVQVVEFPEKSKSKAKLEKGFKEVSEKESDDIDEGDVISOSSPKGKSVDRG  
STISGVSKSKSDSDVKTWTESVDVETGKNDKSKQVKVYIKDKNDGSTERGSD  
TSDQRIDPLRIEKKKTASYIIVKVDGKTVAKEVSDVV"  
/gene="mw1104"  
/705..9580  
/8705..9580  
/gene="mw1104"  
/note="ORFID:MW1104"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAB94969.1"  
/db\_xref="GI:21204272"  
/translation="MTGRIVKSISGVYQVDVNGERENTKPRGLFRKKKFSPPVGDIV  
EPEONINEGYIEOVPERKNEKLRPPVNSNIDTIVYMSAVEPNSFTOLLDRFLYIAHS  
YOLNARIVNKKKTKPIEKOFELNELLKIYENGYEPTEFGNDDBKKIIVENWAGLI  
VLSQSGVGSSTPLNHYRPELNLNENDISKLSNRGKHTIRHVELFERONGHTADTGF  
SALDFDHIDKDELKIDFLLENRIGETKFRNCNHIKEPNVKNVHOLEIGNIAQPRYDH  
YIQLFNEISNRKRY"  
/8581..10225  
/gene="cf1x"  
/8581..10225  
/gene="cf1x"  
/note="ORFID:MW1105  
ribulose-5-phosphate 3-epimerase homolog"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB94970.1"  
/db\_xref="GI:21204273"  
/translation="MKLYPSLLSVDFLDLOHKLRLERAGVDGVHFDVMDGQFVPL  
SIGLPIIDAVRKGTTLPIDVHLMENPEKYIAPFAEHGADMISIHVESTPIHRIQIM  
IKHDKAGVYINPPTISQIEPIIDIVVYVLMVNPFGGOSFIDQCPEKIALGNA  
IKMEROLFNDIEYDGVNDDTAKVCVENGATMLVTGSFFFKQEDYKKVQALG"  
/10232..10873  
/gene="mw1106"  
/10232..10873  
/gene="mw1106"  
/note="ORFID:MW1106"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"

Query Match 99.38; Score 1358.4; DB 1; Length 295350;  
Best Local Similarity 99.68; Pred. No. 5.7e-219;  
Matches 1362; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1 ATGCGATTGAAGGCTTATCAGAACGCTCGCAAGCGACGATGCAAAAATCGCTGTAAG 60  
|||||  
Db 26404 ATGCGATTGAAGGCTTATCAGAACGCTCGCAAGCGACGATGCAAAAATCGCTGTAAG 26463  
|||||  
QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATTAGCGTTATT 120  
|||||  
Db 26464 GGTAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATTAGCGTTACTT 26523  
|||||  
QY 121 GAGGCTGACGTAACTTTAAAGTGGTAAAAAGAAATTTATTAACACAGTATCAGAACGCGA 180  
|||||  
Db 26524 GAGGCTGACGTAACTTTAAAGTGGTAAAAAGAAATTTATTAACACAGTATCAGAACGCGA 26583  
|||||  
QY 181 TTAGGTTCCGANGTAAATGCAATCAATTAACACCGAGCGACAAAGTTATTAAATAGTCAA 240  
|||||  
Db 26584 TTAGGTTCCGANGTAAATGCAATCAATTAACACCGAGCGACAAAGTTATTAAATAGTCAA 26643  
|||||

QY 241 GATCAATTAACGAGCTTGATGGGTGGAGAAATACATCGATTAAATATGTCAAATAAACA 300  
|||||  
Db 26644 GATCAATTAACGAGCTTGATGGGTGGAGAAATACGTCGATTAAATATGTCAAATAAACA 26703  
|||||  
QY 301 CCTACCTGTTGTTATCATGGTGGTTTACAGGTGCTGGTAAACAAACAACCTGCAGGTAA 360  
|||||  
Db 26704 CCTACCTGTTGTTATCATGGTGGTTTACAGGTGCTGGTAAACAAACAACCTGCAGGTAA 26763  
|||||  
QY 361 TTAGCATTATTGATCGCTPAAAAAATAACAACAAACCTATGTTAGTTGACGACGATATT 420  
|||||  
Db 26764 TTAGCATTATTGATCGCTPAAAAAATAACAACAAACCTATGTTAGTTGACGACGATATT 26823  
|||||  
QY 421 TATCGTCCAGCAGCGATAATCAATTAACAACAAGTAGGGAACAACAAATTGATATTCTCTGA 480  
|||||  
Db 26824 TATCGTCCAGCAGCGATAATCAATTAACAACAAGTAGGGAACAACAAATTGATATTCTCTGA 26883  
|||||  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAACATAATGCAATTAACAAATGCT 540  
|||||  
Db 26884 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAACATAATGCAATTAACAAATGCT 26943  
|||||  
QY 541 AAAGAAGAACATTTAGACTTTGTAACTATGATACAGCAGGTGATTTACACATCGATGAA 600  
|||||  
Db 26944 AAAGAAGAACATTTAGACTTTGTAACTATGATACAGCAGGTGATTTACACATCGATGAA 27003  
|||||  
QY 601 GCATTGATCAACGAATTAAGAAGAGTAAAGAAATTCCTAAACCAACAAAGAAATTAATGTTA 660  
|||||  
Db 27004 GCATTGATCAACGAATTAAGAAGAGTAAAGAAATTCCTAAACCAACAAAGAAATTAATGTTA 27063  
|||||  
QY 661 GTTGTGATTCATGAGCGGGPCAGATGCTGCTCAATGTTGCGAGATCTTTTGACGATCAA 720  
|||||  
Db 27064 GTTGTGATTCATGAGCGGGPCAGATGCTGCTCAATGTTGCGAGATCTTTTGACGATCAA 27123  
|||||  
QY 721 CTTGATGTCACAGGTGTTTACCTTAACATAATTAGATGCTGATACAGTGGTGGTGCACCT 780  
|||||  
Db 27124 CTTGATGTCACAGGTGTTTACCTTAACATAATTAGATGCTGATACAGTGGTGGTGCACCT 27183  
|||||  
QY 781 TTATCTATTTCGTTGCTGACACAAACCAATTAATTTGTTGGTATGATGAAAGATTA 840  
|||||  
Db 27184 TTATCTATTTCGTTGCTGACACAAACCAATTAATTTGTTGGTATGATGAAAGATTA 27243  
|||||  
QY 841 GATGGTTTAGACCTATTCCATCTGACGATGATGGCATCGCATATTTTAGTATGGGTGAT 900  
|||||  
Db 27244 GATGGTTTAGACCTATTCCATCTGACGATGATGGCATCGCATATTTTAGTATGGGTGAT 27303  
|||||  
QY 901 GTGTAAAGTTTAAATGAAAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960  
|||||  
Db 27304 GTGTAAAGTTTAAATGAAAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 27363  
|||||  
QY 961 GAGAAAAAGATCGGTGAGTCACTGTTTACTTTAGATGATTTTGTAGAACCACTGATCAG 1020  
|||||  
Db 27364 GAGAAAAAGATCGGTGAGTCACTGTTTACTTTAGATGATTTTGTAGAACCACTGATCAG 27423  
|||||  
QY 1021 GTGAAAAATCTAGGACCACTGGATGATATTATGAAAAATGATCCAGGTATGATATAAATG 1080  
|||||  
Db 27424 GTGAAAAATCTAGGACCACTGGATGATATTATGAAAAATGATCCAGGTATGATATAAATG 27483  
|||||  
QY 1081 AAAGGCTTAGATAAGCTTAAATPATGAGTGAAGCAAAATTTGATCATATTAAAGCGATATC 1140  
|||||  
Db 27484 AAAGGCTTAGATAAGCTTAAATPATGAGTGAAGCAAAATTTGATCATATTAAAGCGATATC 27543  
|||||  
QY 1141 CAGTCAATGACCCCGCTGAAGAAACAATCCAGACACATTCATGATGATCCAGTAAAG 1200  
|||||  
Db 27544 CAGTCAATGACCCCGCTGAAGAAACAATCCAGACACATTCATGATGATCCAGTAAAG 27603  
|||||  
QY 1201 CGTATTGCTAAAGGCTCTGGTCGTTTATTACAAGAAAGTCAATCGTTTGTAGTAAACAAATTT 1260  
|||||  
Db 27604 CGTATTGCTAAAGGCTCTGGTCGTTTATTACAAGAAAGTCAATCGTTTGTAGTAAACAAATTT 27663  
|||||  
QY 1261 AACGATATGAAGAAAAATGATGAACAATTCATCTGCTGCGGTAAAGGTAAAGGTAAAGGTAAA 1320  
|||||  
Db 27664 AACGATATGAAGAAAAATGATGAACAATTCATCTGCTGCGGTAAAGGTAAAGGTAAAGGTAAA 27723  
|||||  
QY 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGCAATTTACCGTTTAA 1368

```
Db 27724 CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTTAA 27771
|||||
E36051 792 bp DNA linear PAT 18-JUN-2001
LOCUS
DEFINITION Signal recognition particle polypeptide and polynucleotide.
ACCESSION E36051
VERSION E36051.1 GI:13022453
KEYWORDS JP 199235183-A/2.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 792)
AUTHORS Michael, T.B.
TITLE Signal recognition particle polypeptide and polynucleotide
JOURNAL Patent: JP 199235183-A 2 31-AUG-1999;
SMITHKLINE BEECHAM CORP
COMMENT OS Unidentified
PN JP 199235183-A/2
PD 31-AUG-1999
PF 03-SEP-1998 JP 199289963
PR 03-SEP-1997 US 60/057890, 05-MAR-1998 US 09/035382 PI
MICHAEL, TERENCE BLACK
PC C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00,
A61K31/00
PC A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/70, A61K35/76,
A61K39/085,
PC A61K39/395, A61K39/395, A61K48/00, C07K14/31, C07K16/12, C12N1/15,
C12N1/19,
PC C12N1/21, C12N5/10, C12P21/02, C12Q1/68, G01N33/15, PC
G01N33/566//
PC C12P21/08, G01N33/577, (C12N15/09, C12R1:445), C12N15/00, C12N5/00,
C12N5/00,
PC (C12N15/00, C12R1:445)
CC Strandedness: Single;
CC Topology: Linear;
FH Key
FT source
FT Location/Qualifiers
FT /organism="Unidentified".
FEATURES
source
1..792
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 296 a 116 c 164 g 216 t
ORIGIN
Query Match 57.8%; Score 790.4; DB 6; Length 792;
Best Local Similarity 99.9%; Pred. No. 4.2e-123;
Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGCTTTGAAGGCTTATCAGAACGCTGCAAGCGAGATGCAAAATATGGGTGAAG 60
Db 1 ATGGCAATTTGAAGGCTTATCAGAACGCTGCAAGCGAGATGCAAAATATGGGTGAAG 60
QY 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATAGCGTTATT 120
Db 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATAGCGTTATT 120
QY 121 GAGCTGACGTAACATTTAAAGTGGTAAAGAAATTTATTAACAGTATCAGAACGCGCA 180
Db 121 GAGCTGACGTAACATTTAAAGTGGTAAAGAAATTTATTAACAGTATCAGAACGCGCA 180
QY 181 TTAGTTCGGATGTAATCAATCATTAACACAGGCGCAAGTATTATAAATAGTTCAA 240
Db 181 TTAGTTCGGATGTAATCAATCATTAACACAGGCGCAAGTATTATAAATAGTTCAA 240
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAATACATCGATTAAATATGTCMAAATAACCA 300
Db 241 GATGAATTAACGAAGTTGATGGTGGAGAAATACATCGATTAAATATGTCMAAATAACCA 300
QY 301 CCTACTGTTGTTATGATGGTTTACANGTGTCTGGTAAACAAACAACACTGCAGTAA 360
|||||
RESULT 5
E36051
LOCUS
DEFINITION Signal recognition particle polypeptide and polynucleotide.
ACCESSION E36051
VERSION E36051.1 GI:13022453
KEYWORDS JP 199235183-A/2.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 792)
AUTHORS Michael, T.B.
TITLE Signal recognition particle polypeptide and polynucleotide
JOURNAL Patent: JP 199235183-A 2 31-AUG-1999;
SMITHKLINE BEECHAM CORP
COMMENT OS Unidentified
PN JP 199235183-A/2
PD 31-AUG-1999
PF 03-SEP-1998 JP 199289963
PR 03-SEP-1997 US 60/057890, 05-MAR-1998 US 09/035382 PI
MICHAEL, TERENCE BLACK
PC C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00,
A61K31/00
PC A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/70, A61K35/76,
A61K39/085,
PC A61K39/395, A61K39/395, A61K48/00, C07K14/31, C07K16/12, C12N1/15,
C12N1/19,
PC C12N1/21, C12N5/10, C12P21/02, C12Q1/68, G01N33/15, PC
G01N33/566//
PC C12P21/08, G01N33/577, (C12N15/09, C12R1:445), C12N15/00, C12N5/00,
C12N5/00,
PC (C12N15/00, C12R1:445)
CC Strandedness: Single;
CC Topology: Linear;
FH Key
FT source
FT Location/Qualifiers
FT /organism="Unidentified".
FEATURES
source
1..792
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 296 a 116 c 164 g 216 t
ORIGIN
Query Match 57.8%; Score 790.4; DB 6; Length 792;
Best Local Similarity 99.9%; Pred. No. 4.2e-123;
Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGCTTTGAAGGCTTATCAGAACGCTGCAAGCGAGATGCAAAATATGGGTGAAG 60
Db 1 ATGGCAATTTGAAGGCTTATCAGAACGCTGCAAGCGAGATGCAAAATATGGGTGAAG 60
QY 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATAGCGTTATT 120
Db 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATAGCGTTATT 120
QY 121 GAGCTGACGTAACATTTAAAGTGGTAAAGAAATTTATTAACAGTATCAGAACGCGCA 180
Db 121 GAGCTGACGTAACATTTAAAGTGGTAAAGAAATTTATTAACAGTATCAGAACGCGCA 180
QY 181 TTAGTTCGGATGTAATCAATCATTAACACAGGCGCAAGTATTATAAATAGTTCAA 240
Db 181 TTAGTTCGGATGTAATCAATCATTAACACAGGCGCAAGTATTATAAATAGTTCAA 240
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAATACATCGATTAAATATGTCMAAATAACCA 300
Db 241 GATGAATTAACGAAGTTGATGGTGGAGAAATACATCGATTAAATATGTCMAAATAACCA 300
QY 301 CCTACTGTTGTTATGATGGTTTACANGTGTCTGGTAAACAAACAACACTGCAGTAA 360
|||||
Db 301 CCTACTGTTGTTATGATGGTTTACANGTGTCTGGTAAACAAACAACACTGCAGTAA 360
|||||
QY 361 TTAGCATTATTGATGCTAAATAATACAAACAAAACCTATGTTAGTTGCAGCAGATATT 420
|||||
Db 361 TTAGCATTATTGATGCTAAATAATACAAACAAAACCTATGTTAGTTGCAGCAGATATT 420
|||||
QY 421 TATGCTCCACAGCAGTAAATCAATTAACAACAGTAGGGAACAAATATGATATTCCTGTA 480
|||||
Db 421 TATGCTCCACAGCAGTAAATCAATTAACAACAGTAGGGAACAAATATGATATTCCTGTA 480
|||||
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACATGCAATTAATAAACAATGCT 540
|||||
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACATGCAATTAATAAACAATGCT 540
|||||
QY 541 AAAGAAGAACATTTAGACCTTTGATATGATATGATACAGCAGGTGCGATTACATCGATGAA 600
|||||
Db 541 AAAGAAGAACATTTAGACCTTTGATATCATTTGATACAGCAGGTGCGATTACATCGATGAA 600
|||||
QY 601 GCATTGATGACGATTAAGAAAGAAAGAAATTTGCTAAACCAACAAAGAAATTAAGTTA 660
|||||
Db 601 GCATTGATGACGATTAAGAAAGAAAGAAATTTGCTAAACCAACAAAGAAATTAAGTTA 660
|||||
QY 661 GTTGTGCGATTCATGACGGGTCAAGATGCTGCTCAATGTTGCAAGATCTTTTTCAGCATCA 720
|||||
Db 661 GTTGTGCGATTCATGACGGGTCAAGATGCTGCTCAATGTTGCAAGATCTTTTTCAGCATCA 720
|||||
QY 721 CTTGATGTCACAGGTTTACCTTAACCTTAATAGATGTTGATACAGTGGTGGTGGAGCT 780
|||||
Db 721 CTTGATGTCACAGGTTTACCTTAACCTTAATAGATGTTGATACAGTGGTGGTGGAGCT 780
|||||
QY 781 TTATCTATTGCT 792
|||||
Db 781 TTATCTATTGCT 792
|||||
RESULT 6
AX413016 319630 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 7 from Patent WO0228891.
ACCESSION AX413016
VERSION AX413016.1 GI:21445474
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 7 11-APR-2002;
Pasteur Institut (FR)
FEATURES
Location/Qualifiers
source
1..319630
/organism="Listeria innocua"
/db_xref="caxon:1642"
BASE COUNT 105207 a 55428 c 66726 g 92263 t 6 others
ORIGIN
Query Match 51.2%; Score 700.8; DB 6; Length 319630;
Best Local Similarity 70.7%; Pred. No. 1e-108;
Matches 933; Conservative 0; Mismatches 387; Indels 0; Gaps 0;
QY 1 ATGGCATTGAAGGCTTATCAGAACGCTGCAAGCGAGATGCAAAATATCGGTGAAG 60
Db 187924 ATGGCATTGAAGGCTTATCAGAACGCTGCAAGCGAGATGCAAAATATCGGTGAAG 187983
QY 61 GGTAACTTACTGAAGCTGATATAAGTAAATGATGCGTGAAGTATGCTGCTAGTATGCTTATT 120
Db 187984 GGTAACTTACTGAAGCTGATATAAGTAAATGATGCGTGAAGTATGCTGCTTACTA 188043
QY 121 GAGGCTGACGTAACATTTAAAGTGGTAAAGAAATTTATTAACAGTATCAGAACGCGCA 180
Db 188044 GAGGCTGACGTAACATTTAAAGTGGTAAAGAAATTTATTAACAGTATCAGAACGCGCA 188103
```



Db 129021 TTTTCTCTAGCGATCAAGTAAGCCAGTAGAAAATCGCGGAACAAAGCATATCGCTAAAGCA 128962  
QY 541 AAAGAAACAACTTTAGACTTTTGAATCATGATACAGCAGGTGCTGATACACATGATGAA 600  
Db 128961 AAAGAAACAACTTTAGACTTTTGAATCATGATACAGCAGGTGCTGATACACATGAA 128902  
QY 601 GCATTGATGACGAATTTAAAGAGATTAAGAAATGCTAAACCAACGAATATATGTTA 660  
Db 128901 ACTCTGATGACGAATTTAAAGAGATTAAGAAATGCTAAACCAACGAATATATGTTA 128842  
QY 661 GTTGCTGATCAATGACGGGTCAAGTGTGCTCAATGTTGAGATCTTTTGGACGATCAA 720  
Db 128841 GTAGTGTATCAATGACTGGGCAAGACCGAGTAATGTTGGCCCAAGACTTCACGACAA 128782  
QY 721 CTTGATGTCACAGGTGTTACCTTAACATAATTAGATGTTGATGATACACGTTGGTGGT 780  
Db 128781 TTGAAATACCGGGTGTGTTATTAACAAATTAGACGGTGTACACGTTGGTGGGACGA 128722  
QY 781 TTATCTATTCGTTGCGTGACACAAACCAATTAATTTGTTGATGATGAGTGAAGATTA 840  
Db 128721 CTTTCCATCGGTTCAGTCACAGGAAACCAATCAATTCGTTGCTACCGGTGAAATATG 128662  
QY 841 GATGGTTTATGAGCTATTCATCCGTAACGATGCGATCACTGATTTTATGATGATGGTAT 900  
Db 128661 GAAGCAATCGAACGCTTCCATCCGATCGTATGCTTCAGAAATCTCGGCATGGGTAT 128602  
QY 901 GTGTTAAGTTAATTGAAAGAGCGCAACAGATGTTGATGATGAAAGAAAGCAAGATTTA 960  
Db 128601 GTACTTTCTTTATGAAAGAGCAACAACTGATGATGATGAGAAATGAAAGATATG 128542  
QY 961 GAGAAAGATCGGTGAGCTATCTTTTACTTTAGATGATTTTATGACACTTGTATCAG 1020  
Db 128541 GAACAAATGAAAGCAACAGCATGACCTTAGATGACTTTTGGACCAATTCGAACAA 128482  
QY 1021 GTGAAATCTAGCCACTGATGATTAATGAAATGATTCAGAGTATGAAATGAAATG 1080  
Db 128481 GTAAACAAATGGACCACTAGATGAATGAAATGATTCAGAGTATGAAATGAAATG 128422  
QY 1081 AAGGGGTAGATGAAGCTTAATGATGATGAAAGCAATGATGATTAATGATGATATAC 1140  
Db 128421 AAGGGGTAGATGAAGCTTAATGATGATGAAAGCAATGATGATTAATGATGATATAC 128362  
QY 1141 CAGTCAATGACGCGCTGAAAGAAACAAATCCAGACACATGAAATGATCACTGATGAA 1200  
Db 128361 AATCCATGACCAAAAGAAAGATTAATCCGACATCAATGATGATGATGATGATGAT 128302  
QY 1201 CGTATGCTAAGGGTCTGGTCTGATTAATGATGATGATGATGATGATGATGATGAT 1260  
Db 128301 CGAATGCTGCTGGAAGCGGACGCCAAATCAAGAAATCAATGCTGCTGCTGCTGCT 128242  
QY 1261 AACGATATGAAGAAATGATGAACAAATTCAGTGGGCTGATGATGATGATGATGATG 1320  
Db 128241 CTTGAAATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 128182

RESULT 8  
AX416333/c  
LOCUS AX416333 2256 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 3324 from Patent WO0228891.  
ACCESSION AX416333  
VERSION AX416333.1 GI:21448790  
KEYWORDS Listeria monocytogenes ATCC 19115.  
SOURCE Listeria monocytogenes ATCC 19115  
ORGANISM Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
REFERENCE 1  
AUTHORS Glaser, P. and Kunst, F.  
TITLE Listeria innocua, genome and applications  
JOURNAL Patent: WO 0228891-A 3324 11-APR-2002;  
Pasteur Institut (FR)  
FEATURES  
Location/Qualifiers  
i. .2256  
/organism="Listeria monocytogenes ATCC 19115"

BASE COUNT 533 a 471 c 420 g 832 t  
ORIGIN  
Query Match 51.0%; Score 697.6; DB 6; Length 2256;  
Best Local Similarity 70.9%; Pred. No. 1.2e-107;  
Matches 931; Conservative 0; Mismatches 389; Indels 0; Gaps 0;  
QY 1 ATGGCATTTGAGGCTTATCAGAACCCCTGCAAGCGACGATGCAAAAATGCGTGGTAAAG 60  
Db 1666 ATGGCATTTGAGGCTTATCAGAACCCCTGCAAGCGACGATGCAAAAATGCGTGGTAAAG 1607  
QY 61 GGTAACTTACTGAGCTGATTAATAAGATTAATGATGCGTGAAGTAAAGATTAATGCGTTATT 120  
Db 1606 GGTAACTTACTGAGCTGATTAATAAGATTAATGATGCGTGAAGTAAAGATTAATGCGTTATT 1547  
QY 121 GAGGCTGACGTAAACTTTAAAGGTGTAAGAAATTTATTAACACAGTATCAAGACGCGCA 180  
Db 1546 GAGGCTGACGTAAACTTTAAAGGTGTAAGAAATTTATTAACACAGTATCAAGACGCGCT 1487  
QY 181 TTAGGTTCCGATGTAATGCAATCATTAACCCAGGGCAACAAGTTATTATAATAGTTCAA 240  
Db 1486 GTGCGGCGGAGCTTATGAAAGCCCTAACACCCGCCCAACAAGTTATTATAATAGTTCAA 1427  
QY 241 GATGAATTAACGAAGTTCATGCGTGGGAGAAATATCATGATTAATATGTCAAATTAACCA 300  
Db 1426 GATGAATTAACGAAGTTCATGCGTGGGAGAAATATCATGATTAATATGTCAAATTAACCA 1367  
QY 301 CTTACTGTTGTTATGATGTTGTTTACAAAGTGTGTTTACAAAGTGTGTTTACAAAGTGTGTT 360  
Db 1366 CCAACCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1307  
QY 361 TTAGCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
Db 1306 CTTGCTTAATTTATGCAAAAATTAATTAACCGTAAACCAATTAATTAACCGTAAACCAAT 1247  
QY 421 TATGCTCCAGCAGCGATTAATCAATTAACAGTAGGAGAAACAAATGATGATGATGATGAT 480  
Db 1246 TACCGACCTGCGGCAATCAACAAATTAAGAACACTTGGCAAGCAATTAAGATGATGATGAT 1187  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCCAACAATTTGTAACATAATGATGATGATGATGAT 540  
Db 1186 TTTTCTCTAGGACCAAGTAAGCCCGGTAGAAATCGCCAAACAAGCTATGCTAAAGCA 1127  
QY 541 AAAGAGACATTTAGACTTTTGAATCATGATACAGCAGGTGCTGATACACATGATGATGAA 600  
Db 1126 AAAGAGACATTTAGACTTTTGAATCATGATACAGCAGGTGCTGATACACATGATGATGAA 1067  
QY 601 GCATTGATGACGAATTTAAAGAGATTAAGAAATGCTAAACCAACGAATATATGTTA 660  
Db 1066 ACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1007  
QY 661 GTTGCTGATCAATGACGGTCAAGTGTGCTGATGATGATGATGATGATGATGATGATGAT 720  
Db 1006 GTTGCTGATCAATGACGGTCAAGTGTGCTGATGATGATGATGATGATGATGATGATGAT 947  
QY 721 CTTGATGTCACAGGTGTTTACTTAATTAATGATGATGATGATGATGATGATGATGATGAT 780  
Db 946 TTAGAAATTTAGAGTGTGCTGCTGATACCAATTTAGACGGTGTATACACATGATGATGAT 887  
QY 781 TTATCTTATTCGTTGCGGTGACACAAAACCAATTAATTTGTTGATGATGATGATGATGAT 840  
Db 886 CTTTGGATTCGTTGCGGTGACACAAAACCAATTAATTTGTTGATGATGATGATGATGATGAT 827  
QY 841 GATGGTTTATGAGCTATTCATCTGCAACGATGATGATGATGATGATGATGATGATGATGAT 900  
Db 826 GAAAGCTCTAGAAACTTTTCCATCCAGATGCTATGCTTCCAGAAATTTCTCGGCATGGGCGAT 767  
QY 901 GTGTTAAGTTTAAATGAAAGCGCAACAGATGTTGATGATGATGATGATGATGATGATGAT 960  
Db 766 GTGCTTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 707  
QY 961 GAGAAAGATGCGTGTGATCATCGTTTACTTTAGATGATTTTATGAAACAACTTGTATCAG 1020

```

Db 706 GAACAAAAATGAAGAACACACAGCATGACGCTAGATGACTTCTAGAACAGTTACAACAA 647
QY 1021 GTGAAAATCTAGACCACCTGATGATATATGAAAATGATTCACGATGATGAATAAATG 1080
Db 646 GTAAACAAATGGGACCATAGATGACTTAAATGATGCGAGGCGCAACAAATG 587
QY 1081 AAAGGGCTAGATAAGCTTAATATGATGAAAAGCAAAATTCATCATATTAAGCGATTATC 1140
Db 586 AAAGGCTCGACACATGACGCTGATGATAAACAATAGTCAATATGAGCAATATC 527
QY 1141 CAGTCAATGACGCGCTGAAGAAACAATCCAGACATATGAATATCATGCTAAAG 1200
Db 526 AAGTCAATGACCAAAATGAAAAGCAATCCAGATATCATCAATCAAGTAGAAGAAA 467
QY 1201 CGATGCTAAAGGCTGCTGCTTCATTACAGAAAGTCAATGATGATGAACAATTT 1260
Db 466 CGATGCTGACGCTGACGCGAGCTCGGTTCAAGAAATTAACCGTCTCTCAAAACAATTT 407
QY 1261 AAGCATATGAAGAAATGATGAACAATTCACCTGCTGCGGTAAGGTAAAAAGGTAAA 1320
Db 406 GGTGAATGAAGAAATGATGAAGCAATGACTGCTGCGGCAAGGCAAAAGGCAAA 347

RESULT 9
D14356
LOCUS
DEFINITION Bacillus subtilis orf1, f1f, rpsP genes for ORF1, F1f and 30S
                ribosomal protein S16, complete cds.
ACCESSION D14356
VERSION D14356.1
KEYWORDS rpsP; orf1; 30S ribosomal protein S16; f1f.
SOURCE Bacillus subtilis (strain:168) DNA.
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 2739)
AUTHORS Honda,K., Nakamura,K., Nishiguchi,M. and Yamane,K.
TITLE Cloning and characterization of a Bacillus subtilis gene encoding a
        homolog of the 54-kilodalton subunit of mammalian signal
        recognition particle and Escherichia coli F1f
JOURNAL J. Bacteriol. 175 (15), 4885-4894 (1993)
MEDLINE 93328695
REFERENCE 2 (bases 1 to 2739)
AUTHORS Yamane,K.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1993) Kunio Yamane, Inst. Biological Sciences,
        Tsukuba University, Tsukuba, Ibaraki 305, Japan (Tel:0298-53-6419,
        Fax:0298-53-6006)
FEATURES
    source
        1..2739
            /organism="Bacillus subtilis"
            /strain="168"
            /db_xref="taxon:1423"
        292..297
            /evidence=experimental
        315..320
            /evidence=experimental
        365..697
            /gene="orf1"
        365..697
            /gene="orf1"
        /codon_start=1
        /transl_table=11
        /product="ORF1"
        /protein_id="BAA22221.1"
        /db_xref="GI:242968"
        /translation="MSLEKTRNMYLDFYQSLTSSKQSYMSLYLDDFSLGEIAEE
        YEVSRQAVVDNIKRTKTEAMLEQYEKLLLLKKFQERKEMFNKLKELASGSKSEETAL
        ITALKELD"
        711..2051
            /gene="orf1"
        711..2051
            /gene="orf1"
            /gene="f1f"

```

```

/codon_start=1
/transl_table=11
/product="F1f"
/protein_id="BAA21691.1"
/db_xref="GI:2309080"
/translation="MAPEGLADRLQOTISKIRKGKVSQDVVKEMREVEFLALLEADY
NFVKYKDFVKYSEFVANGODVMSLPTGGQVILKVQBELTBLMGESBESKTAFAKRPPT
VIMVGLQAGAKTTSKLANLRLKKHNKPMVAADIRFPAALQKLETIGKOLDMPV
FSLGDDVSPEIAKQAEKEEHYDIVILDTAGRLHDHDELDLNNKELANPEEI
FLVDSMGQDPAVNVAKSFNEQLGLTGVVLTIKDGTGRGAALSIRAVTNTPIKFAGL
GEKLDALFFHPFPMASRIILMGVDLTLLIEKAQASVDEDEKAKLEQKMRMTSLDDF
LEQGVNRMGPLDELLQMPGAKMGLNKIQTVDKOLNHNVEAIKSMVTLEKEQPD
IINASRRKRIAGSGTSTVQEVNRLKQFDEMKNMKMQTNNSKKGKFKLPM"
-35_signal 2091..2096
-10_signal 2117..2122
gene 2157..2429
        /gene="rpsp"
        2157..2429
            /gene="rpsp"
            /codon_start=1
            /transl_table=11
            /product="30S ribosomal protein S16"
            /protein_id="BAA21692.1"
            /db_xref="GI:2309081"
            /translation="MAVKIRLKRMAKSKSPFYRVIVADSRPRDGRFIETVGTYNPVA
            KPAEYKIDEEALALQKLGAKPSDVTYRNLFSQGIEMKXFNKAKGK"
BASE COUNT 887 a 527 c 648 g 677 t
ORIGIN
Query Match 45.5%; Score 622.8; DB 1; Length 2739;
Best Local Similarity 67.1%; Pred. No. 3.9e-95;
Matches 882; Conservative 0; Mismatches 432; Indels 0; Gaps 0;
QY 1 ATGGCATTTGAAGCTTATCAGAACGCTGCAACGACGACGATGCAAAAATGCGTGAAG 60
Db 711 ATGGCATTTGAAGGANTTAGCGACCGACTGCGACGACGAGTTCCTAAATTCGCGGAAA 770
QY 61 GGTAACTTACTGAAGCTGATTAAGATATATGATGCGTGAAGTAGATTACGCTATT 120
Db 771 GGGAAAGTCAGCAACAAGATGTAAGAGATGATGCGTGAAGTTCGCTTTCGCTGCT 830
QY 121 GAGGTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACATATCAGAACGCA 180
Db 831 GAGGTGACGTAACTTTAAAGTAGTCAAGATTTTGTCAAAAAGTAGTAGAGCGCT 890
QY 181 TTAGTTCGCGATGATGCAATTAACACGAGGCAACAAGCTTTATTAATAAGTTCAA 240
Db 891 GTAGSCCAAGACGTCATGAAAAGTCTGACGCCGCCAGCAGCTCATTAAGTTGTTCAA 950
QY 241 GATGAATTAAGAGTGTGATGGTGGAGAAAATACATCGATTATATGTCAAATAACCA 300
Db 951 GAGGACGTGACTGAGCTGATGGGCGGCAAGAGAGCAAAATCCCGTCGCAAAAAGCGG 1010
QY 301 CCTACTGTTGTATGATGTTGTTTACAAGGTGCTGTTAAAAACACACTCAGGTAAA 360
Db 1011 CGGACTGTTATGATGTTGCTGCTCCAGGTGCGGTGTAACGACGACGAGGTAAG 1070
QY 361 TTAGCATTTATGATGCGTAAATAATACAAACAACTATGTTAGTTGACGAGATATT 420
Db 1071 CTTGCAATCTGCTGCGCAAAAAGCATTAATGCAAAACCGCATGCTGTTGCTGCGCAT 1130
QY 421 TATGTCACGACGATTAATCAATTAACACATAGGGAACAATGATATTCCTGTA 480
Db 1131 TACGCGCCGCGCCCAATTAAGCAGCTGGAACACTCGGCAACAGCTTGATATGCTGTT 1190
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAATTTGTAATGTCATTAACATGCT 540
Db 1191 TTCCTCTTGGCGATCAGTCAGTCTCTGTAGAAATAGCTTAACAGGCTATTGAGAAAGCC 1250
QY 541 AAGGAAGAACATTTAGACTTTTAAATCATTTGATACAGAGGTCGATTACATCGATGAA 600
Db 1251 AAGGAAGAACATTTAGCTACGTCATTTTGGATACGCGAGGCGCTTGCATATCGACCAT 1310

```



```

QY 601 GCATTGATGACGCAATTAAGAGAGTAAGAAATTCGTAACCAACAAAGAAATTTAGTTA 660
Db 1311 GAACATGATGATGATGACGACCAACGTCGAAGAAATTCGGAATCGGAAGAAATTTTCCTG 1370
QY 661 GTTGTCGATCAATCAAGCGGTCAGATGCTGTCAATGTGTCAGAACTCTTTGACGATCAA 720
Db 1371 GTTGTCGATCAATCAAGCGGTCAGGAGCGTGTGGAATGTGCCAAAGAGCTTTAATGAAC 1430
QY 721 CTGTGATGTCACAGGTTTACCTTAACATAATAGATGTGATGATGATGATGATGATGATG 780
Db 1431 CTCGGTTAAACCGGTGTGTGTGTAAGCTGGATGAGACACACGCGCGGCGGCTCG 1490
QY 781 TTATCTATTCCTGTCGTCACACAAACCAATTAATTTGTTGATGATGATGATGATGATGAT 840
Db 1491 CTTTCTATTCGCGTGTGCACACACCGCCCAATTAAGTTTGAGTTTGCGGCGCAAGAGCTT 1550
QY 841 GATGGTTTAGAGCTATTCCATCCTGCAACGTGATGATGATGATGATGATGATGATGATGAT 900
Db 1551 GATGCGTTAGAGCGGTTCCATCCTGCAACGATGATGATGATGATGATGATGATGATGATG 1610
QY 901 GTGTTAAGTTTAATGAAAGCGCAACAAAGATGTGATGATGATGATGATGATGATGATGATG 960
Db 1611 GTGCTGACATTTGATGAAAGCGCAACAAAGATGTGATGATGATGATGATGATGATGATGATG 1670
QY 961 GAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
Db 1671 GACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1730
QY 1021 GTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
Db 1731 GTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1790
QY 1081 AAGAGGCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
Db 1791 AAGGCGCTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1850
QY 1141 CAGTCAATGAGCGGCTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
Db 1851 AATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1910
QY 1201 CGTATGCTAAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
Db 1911 CGGATGCAAAAGAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1970
QY 1261 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1314
Db 1971 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2024

RESULT 10
BSUB0009
LOCUS      208780 bp      DNA      linear      BCT 26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 9 of 21): from 1598421
            to 1807200.
ACCESSION  Z99112 AL009126
VERSION    299112.1 GI:2633902
KEYWORDS   .
SOURCE     Bacillus subtilis.
ORGANISM   Bacillus subtilis.
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 208780)
Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
Azevedo,V., Bertsch,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
Cartier,N.M., Choi,S.K., Codani,J.J., Conwell,R.F., Cummings,N.J.,
Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
Emmerson,P.T., Eniian,K.D., Errington,J., Fabret,C., Ferrari,B.,
Faulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
Gallener,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
Grandi,G., Guisepi,G., Guy,B.J., Haga,K., Halech,J., Harwood,C.R.,
Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,

```

```

Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
Larinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
Masuda,S., Maue,C., Medigue,C., Medina,N., Mellado,R.P.,
Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
Ogawa,K., Ogihara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
Purnelle,C., Rocha,E., Roche,B., Rose,M., Reynolds,S., Rieger,M.,
Rivolta,C., Rocca,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenecker,T.,
Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
Danchin,A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
9384377
2 (bases 1 to 208780)
Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr.
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
Location/Qualifiers
1. 208780
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
258..4588
/genes="bpr"
258..4559
/genes="bpr"
/EC_number="3.4.21.-"
/notes="alternate gene name: bpf"
/codon_start=1
/transl_table=11
/product="Bacillopeptidase F"
/protein_id="CAB13404.1"
/db_xref="GI:2633903"
/db_xref="SWISS-PROT:P16397"
/translation="MRKTKNRLISSVLSTVVISLPPGAGASSKVTSPSVKKELQ
QKSAVSVSLKTKADESCQDVLKYLNTQKDKGNADQIHSYIVNGIAVHASKEVMEKY
VQPEKVEKLVNPKRQLFKSSPFNMKKAQKATKATGVEWNVQDIDAPKALGYDG
TGTVVASIDTGVWVHPALKERYGNPNPENPENNNWYDAVEASPYDDLAHGT
HVTGTMGSEPDGTNOIGVAPGAKTAVKAFSEDGDTADILEAGEWVLPKDAENP
HPMAPVYVNNSSGGGGLDEWYDMVNAARAADIFFEFSSAGNTDLFIPGPGSTANP
ANYPEFAFGATDINKLADFSLOQSPYDEIKPRIASPGVNIERSPGTQVDEGDWG
TSAGPHYSVAALLKQANASLSYDEMDILISTAEPLTDTSPDSFNGYGHGLVNA
FDVASVDGKAGQVSGVDDQEPVYQHEKVTEAYEGGSLPLTFADNVSVTS
VKLSKLDGQEWETITAKRISGDHLKGTQAEIPDIKTKLSYKMHDPFGVSSD
VIDVTVPKTFITAGYKQDFETAPGQWVASGTNNWVSPGTPNTAASGEKYVGNLT
GYNASANNLVMPPIKAPDGSGLFLOKSWHLEDPDFYGVFVLPPEKQWQAGV
NKGTSWPTDEEDLSAYKQNTQVFNQSDSIAKRGWYIDVVLVLSKSAKTKVK
NKLGVKEPGQKKPKPKPKPSANTAVKHQKATQPVLLPKAQVSVVETGSIY
SDSGTGYTLKHKAGDTLMAEAYGQSKTKVSLKTDQTOANFLTNEMKGTUKGT
VINKTGVETGASVYVEADAEPAMTNGEYMLAEYBGATIKVAPAGYSDSFS
VELGDKTKETALKPFYFCEIATYDDGTANANSYFAAGNWAIVKMTLADGDKGML
TGLGPRFWPTFPDPGTEEKVEYDATGKGCGKTIAGTFNAELRQVTKVDLS
SKGLMVKDFVLVYIQSKPDPYSPGLAMDETQNSGRWQYIDGKQWPGDKAGNMI
RALYDEARVEELISPDKSYTKNDKSVYVKNASPGTVHYHNGEAEGETKAADGT
FHAGILNKEUTASTADNGTDTASSPITVLDQKPELIDNPDKGKTKETTL
TVKAGVSDNNKDKVKNGKATGADGYSARILLNGRKEIKVATDLAGNKTTKTY
IDVNFDPKVISGLIPGDKNKAGESVKIAFSAAEDLDATFTIRMPITNARASQVAT
ELPFRISPGRYEGYWTATSSIKAKGAKVEVIVRDYDQNETRKTANGKLNMTNEN"

```

```
terminator
4571..4580
/gene="bpr"
4754..5683
/gene="spoiIGA"
4754..5683
/gene="spoiIGA"
/function="processing of pro-sigma-E (SpoIIGB) to active
sigma-E (stage II sporulation)"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="CAB13405.1"
/db_xref="GI:2633904"
/db_xref="SWISS-PROT:P13801"
/translation="MKIYLDVILWLNFCFDALLLLAFILKHKVKRLVGGAFIGS
SVLLMPTPRSPVHPAGKIAFSVIVVVFGEKRRFFQNLFSFPAFLMGGI
IGASLGSISIVQMGVMTNCTGDPISLWLFVGGFPALWFPSKRRIEDIEKNIQ
YERVSQADLGSQTLHVRLGIDSGNQLIDPLTIPWMIYIDKLEPIFPTAETMIR
NTPLEAIEQDDSFRLDPMKRLIPYRGVQGNQFLLCVKPDHVTIMTKEEMISADKC
LIGSTTKLSADGEFDALIHPRKMLSGKAVKVS"
5746..6465
/gene="sigE"
5746..6465
/gene="sigE"
/function="early mother cell-specific gene expression"
/note="alternate gene name: spoIIGB"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma-29
factor (sigma-E)"
/protein_id="CAB13406.1"
/db_xref="GI:2633905"
/db_xref="SWISS-PROT:P06222"
/translation="MKKLRLTHLWYKLMKLGKSDVEYVIGSSEALPPPLSKDEE
QVLLMKLPNGQDQARAILLERILRLVVIARKFENGGINEDLSIGTIGLIKAVNPF
NPKKILATYASRCIENEILMLRNKIRSEVFEPLNDWDGNELLSDVLGTD
DDITKDIEANVDKLLIKKALQLNREKQINELRFLVGLVEEEKTKQVADMMGTSQS
YISLRKRIIKRLKRFNKV"
6605..7387
/gene="sigG"
6605..7387
/gene="sigG"
/function="late forespore-specific gene expression"
/note="alternate gene name: spoIIIG"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma factor
(sigma-G)"
/protein_id="CAB13407.1"
/db_xref="GI:2633906"
/db_xref="SWISS-PROT:P19940"
/translation="MSRNKVEICGVDTSKLPVLKNEEMKRLFQLQDEGDSAREKL
NGNRLVLSVIOFRNNGEYVDLQVGCIGLMSIDNFDLSHNKFSFYAVPMIGE
TERVLNNRIRVRSRLRIAYKALQVRELISETKEPTAEIAKVLVPEHEIYFA
LKAIDPFQVFEPIYNDGGDPYVMDQSDKRDNTDSOWIEELAKEGMRRLNDRKMI
LRKIDPFQVMEYAEIGISQAQVSRLEKAAIKOMANKNIHQ"
7535..8496
/gene="ylmA"
7535..8496
/gene="ylmA"
/function="unknown"
/note="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB13408.1"
/db_xref="GI:2633907"
/db_xref="SPREMBL:O31723"
/translation="MTIQDNLVSLKRNKGWILKDHWKVEKENWVLYLNGAGKNTAL
LNMLGSYFTSPGEMVIGHEFKTELKRLKRLIGLVSAALOOKLYPADSAFELALS
GAYASIGYFTPPKSTREKAIGLLEDGAIEYADRYETLSQGEKORALIFALMADP
ELLILDPEVLGDFIAREKLLDITFIANKENAPSILYVTHAEELIPYFDKALLLKQ
GVYFSGSEIKEMUTDQLSADFDTPIHVNQDRPFLTRAEPIINA"
8472..8496
terminator

/gene="ylmA"
8531..9811
/gene="ylmB"
8531..9811
/gene="ylmB"
/function="unknown"
/note="similar to acetylornithine deacetylase"
/codon_start=1
/transl_table=11
/protein_id="CAB13409.1"
/db_xref="GI:2633908"
/db_xref="SPREMBL:O31724"
/translation="MDQIYSIQKVEHKEELIQAKTLISYQTAPPARNTGSIQS
WDPPVPLKNGLLIGRCASDMKGMACVLFVKLIRASIELPDLLIOSVIGEEVGE
AGTLECCRGVHADPAIVADTSDMHIOQGGVITGWIEIKSSOTFHDGRRNRMIHAG
GTFGSALERKWTAKIAGLELBEHWSIMKSYPGFKPTNTINFAVEGGHAAFTADE
CLMTWTFYFNETHQVAAEIEDVYNRLSDIWLNRNPFVKWGGSSMIIDRGEIF
PALEYDPGPGVLAATASHOKVRKRCFIIYDSQSVTDGGWLYDAGIPCVIYGGDLHN
AHSVNEKYSIBOLVETVKIILDFIISWCSRKEQ"
9894..10139
/gene="ylmC"
9894..10139
/gene="ylmC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB13410.1"
/db_xref="GI:2633909"
/db_xref="SPREMBL:O31725"
/translation="MISIEFQVKDVNVNNGKLGISGIDIDINVTGKIQAIIILGGN

Query Match 45.5%; Score 622.8; DB 1; Length 208780;
Best Local Similarity 67.1%; Pred. No. 1.3e-95;
Matches 882; Conservative 0; Mismatches 432; Indels 0; Gaps 0;

QY 1 ATGGCAATTTCAGAGCTTATCAGACGCGCTCGACGCGATGCAAAAAATCGGTGTAAG 60
DB 73153 ATGGCAATTTCAGAGATTAGCGCGACTGCGACGACAGATTCCTAAATCCGCGGAAA 73212

QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAAGATTAGCGTTATT 120
DB 73213 GGGAACTCAGCGCAACAAGATGATAAGAGAGATGATCGGTGAGTCCGCTTCCTGCTT 73272

QY 121 GAGCTGACGCTAACTTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGCGCA 180
DB 73273 GAGCTGACGCTTAACTTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGCGCT 73332

QY 181 TTAGGTTCCGATGATTAATGCAATCATTAACACGAGGCAACAAGTTATTAAATAATAGTTCAA 240
DB 73333 GTAGCCCAAGACGCTCATGAAAGTCTGACGCCGCCGACGAGTCAATTAAGTTGTTCAA 73392

QY 241 GATGAATTAACGAAGTGTATGGGTGGAGAAATAATACATCGATTAATATGTCATAAATAACCA 300
DB 73393 GAGGAAGTGAAGTGTATGGGTGGAGAAATAATACATCGATTAATATGTCATAAATAACCA 73452

QY 301 CCTACTCTGTTATGATGTTGTTTACAAAGTGTGTTAAACACACAACTGCGAGTAAA 360
DB 73453 CCGACTGTTTATGATGTTGTTTCCAGGTTGCGGTGTTAAACACACAACTGCGAGTAAA 73512

QY 361 TTAGCATTTATGATGTTAAATAATACAAACAAAACCTATTGTTAGTTCACAGCATATT 420
DB 73513 CTTGCGAATCTGTCGCAAAAAGCATATAATCGCAACCGATGCTGTTGTCGCGATATT 73572

QY 421 TATCGTCCACAGCATTAATCAATTAACACAGTAGGGAACAATAATGATTAATTCCTGTA 480
DB 73573 TACGCCCAAGCCGCAATTAAGCAGTGGAAACACTCGGCAACACAGCTTGATATGCTGTT 73632

QY 481 TACAGTGAAGAGATCAAGTAAAGCCACCAAAATGTAACTAATGCAATTAACACATGCT 540
DB 73633 TTCTCTCTTGGGATCAGGTCAGTCTCTAGAAATAGCTAAACACAGGCTATTGAGAAAGCC 73692
```



```
/translation="MSLNKQIKSEHDHNLKLVSNQVTVIFALNLFALILEFIFGV
LNFSTAILSDAHTDQAVAGLAWFQKSKREDKHSFGQFORSLLGASLTGVIL
ITGSLTIVPEAVPREFNPQVEATGMGLAIFAIVANGFGAWLLARGSSNESILNH
ALEDYGLHGLVIVSYVHFVXWYLDPLLSILIALFILSKAIPKFWGILKILLESYP
EDIDKRALQELQEPVLAHTVQLIWSIDGQNAAMIHIVIPENQDFSDAKIAVRKL
LESEKVSASALEDDTDEHKHQVEI"
gene      complement(2940. .4181)
          /gene="proa"
          /note="I0116"
          /complement(2940. .4181)
          /gene="proa"
          /note="EVIDENCE BY HOMOLGY BIO01.05 Glutamate family.
SECOND STEP IN PROLINE BIOSYNTHESIS PATHWAY"
          /codon_start=1
          /transl_table=11
          /product="gamma-glutamyl phosphate reductase (EC
1.2.1.41)"
          /protein_id="AAK05706.1"
          /db_xref="GI:12724615"
          /translation="MIEELGLVKKASKEVAKLSTADKNVFLQNLADSLIENTDRIIS
ENEKDLANALEHGISETMDORLNAQRIQSDMATGLROVAELPDPIROVLOGFTNLDG
LKLTKRVPGLTGVGMIFESRPNTIDAFSLCFKTSVLLAGSGDALIYSNVLVELIK
ENLLSAKTIDGAVELSDTSHAERQMQRADFDVLIPRGSARLNRKREKATVPVI
ETVGNCITFVDESADLEMATKIVINAKTORPSCVCAASLVHAKIADIEFUPKLENE
INKVHEIFRADRALKVLGAGIPATDDDFGEFDLYILSVKVDNIEAIBHINTYS
SRHSSEIVTHDYFNAQRFQDEIDAAAAYVYNASTRTDGFVEGLGAEIGISQKLEHARG
PMGLEALTSTKYLIDGTQIR"
gene      complement(4199. .5011)
          /gene="proB"
          /note="I0117"
          /complement(4199. .5011)
          /gene="proB"
          /note="EVIDENCE BY HOMOLGY BIO01.05 Glutamate family.
FIRST STEP IN PROLINE BIOSYNTHESIS PATHWAY"
          /codon_start=1
          /transl_table=11
          /product="glutamate 5-kinase (EC 2.7.2.11)"
          /protein_id="AAK05707.1"
          /db_xref="GI:12724616"
          /translation="MIMTRKNILSVKRIYVKIGTSSLLPNKINLSNIDELAFVLSO
LNKNGVEVILVTSAGVGLNVLGIDKRPKGIADQOALASIGQVFLMSLYTQMFRRYS
QKVSQLLTRDVTDFPYSRENAENALNLLGIGIPIINENDAIJVDMDHDTKFGDN
DKLGAIVSKNADLIMLSDIGLFDKPNISYDDAKIFNEIHEIDELROMAGGAGS
RFGTGMSTKLAAQILFENGQEWLNTGERIREIKETIEIGREIGTYFHQKS"
gene      complement(5141. .6613)
          /gene="yqfA"
          /note="I52568"
          /complement(5141. .6613)
          /gene="yqfA"
          /note="EVIDENCE BY HOMOLGY BIO15.02 HYPOTHETICAL.
Conserved"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AAK05708.1"
          /db_xref="GI:12724617"
          /translation="MTNMNKSITIERQKRIVEILTLHEGGMFDKAKIFNEFDGVD
VSPITGAERALKOGLNPSSEIOLCNVHAAPKGAITDIHKSSLEENTPGHPHTHMKL
ENQVITSLNDEIRWFSKIEKGDSWKRELLAVIDLYNDIKHYAKESILFSPNEK
IGTAPKVMVGDDSVREMIKEVIYFLKTDALNPLREMLNLLTIEEMIFKEEA
IMTPNCLDVFSLKQWEQTEESKEISGYATFAEPGLWKAKSEIEKEKEHPQRIAAE
SAKAMTEALAKESGLENSKPKRVKQYDWKVTSQGVVPLTGMHLNELTLFNVLP
VDLSFQKEDIVRRPSSGGERIFRPAKSVIGRVDPHPKSNVAVKILBDRKSGPRD
QADFWDLHRFNKITYIRYFAMRDEETGEYGLGLEVSODITISIQNLVGEKRLDGHEKT
TSEYENQVAKSIKSTDGAVNGEMPDPVFKMLAEKTKNS"
gene      complement(6613. .6852)
          /gene="yqfB"
          /note="I54040"
          /complement(6613. .6852)
          /gene="yqfB"
          /note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
          /codon_start=1
          /transl_table=11
```

```
/product="UNKNOWN PROTEIN"
/protein_id="AAK05709.1"
/db_xref="GI:12724618"
/translation="MIQIDNNAKLYDLATEHPEIIDLMDDLGFHEIKPMGMLQTAGRM
ATIPGAKKHIDWDKIVIAFAPAGFEFSNQVEE"
gene      complement(6979. .7305)
          /gene="yqfC"
          /note="I54406"
          /complement(6979. .7305)
          /gene="yqfC"
          /note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
          /codon_start=1
          /transl_table=11
          /product="UNKNOWN PROTEIN"
          /protein_id="AAK05710.1"
          /db_xref="GI:12724619"
          /translation="MSSFEFISLTIIIGCAIVTWISRVLPFLLKKMSLPQIVVEYLSF
VPVIMSALWISNFIQHLGELFSVNNLLASIPTVLAAILTKNLLVIVLVGVSFA
FIQIIP"
gene      complement(7292. .7999)
          /gene="yqfD"
          /note="I54719"
          /complement(7292. .7999)
          /gene="yqfD"
          /note="EVIDENCE BY HOMOLGY BIO13.02 TRANSPORT AND BINDING
PROTEINS. Amino acids, peptides and amines"
          /codon_start=1
          /transl_table=11
          /product="amino acid permease"
          /protein_id="AAK05711.1"
          /db_xref="GI:12724620"
          /translation="MNSELTFKQGLKDTPTVFGYIGIAGFMIGHSEGSFVWVILL
LSLIVYAGSAQFIMVSLMATHPSIMSVLSPVNSRIILMSMTASYPKNESLLKNI
LLGLTLTDESAFLGMKNKONTGKLNFSNELLAYLVWALSAIGALLGNLANP
EKLGIGFVAINFIFGLLYLQLISDKTLGLMLQLVMVGITLVIFYFGLIFLPSNLLVLF
VTIIACALGVGVKRVFF"
Query Match      41.4%; Score 567; DB 1; Length 11023;
Best Local Similarity 64.7%; Pred. No. 6.3e-86;
Matches 843; Conservative 0; Mismatches 460; Indels 0; Gaps 0;

QY 1 ATGGCATTGGAAGCCTTATCAGAACGCTGCAAGCAGCATGCAAAAATGCGTGGTAAG 60
Db 10954 ATGGCATTGGAAGCCTTAACTGAACGCTTACAAAACGCTTTTAAAAATCTACGTGGAAG 10895

QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGTCGCTGAAGTAGAATAGCGTTATT 120
Db 10894 AAAAAATTACTGAGACAGAGCTCAGAGAAATTACAAAAGAAATCCGAGTTCCTCT 10835

QY 121 GAGGCTGACGTAAACTTTAAAGTGGTAAAGAATTTATTAACACAGTATCAGAACGCA 180
Db 10834 GAAGCCGAGCTTGCTCTCCAGTGTGCAAGAGTTTATTAAGGCAATCCGTGAACGAGCT 10775

QY 181 TTAGTTCCTGATGTAAATGCAATTAACACAGGCAACAGTAATTAATAATGTTCAA 240
Db 10774 GTCGCTGTTGAAGTATCAGAAAGCTTAACCCAGCTTAATAAATTTGTTGAT 10715

QY 241 GATGAATTAACGAAGTGTAGGCTGGAGAAAATACATCGATTAATATGTCAAATAAACA 300
Db 10714 GAAGAATTAAACAGCTATTTTGGGTGGTGGCAGAGTGAATTAATCTCTCTCTCTCT 10655

QY 301 CCTACTGTGTTATGATGGTGTGTTTACAAAGTGTGTTTACAAAACAACTCAGAGTAAA 360
Db 10654 CCAACAAATTAATGATGTTGGTGTGCAAGGGCAGSTAAACACACTTTTCTGTTAAA 10595

QY 361 TTAGCATTATTGATGCGTAAAAAATACAAACAAAACCTATGTAGTTGCACAGATATT 420
Db 10594 CTTGCCAAAACCTCAAAAGAGAGCAAAATGTCGCCCTTTAATGATTCGACGCCAGCT 10535

QY 421 TATCTCCACAGCAGATAATCAATCAATTAACACAGTAGGGAAACAAATTTGATTCCTCT 480
Db 10534 TATCTCCACAGCAGGATTGACCAATTTAAAAACCTTAGGTGAACAGCTTGAAATTCCTCTC 10475
```



(E-mail: takamih@jamstec.go.jp,  
URL: http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Research.html,  
Tel: 81-468-67-3895, Fax: 81-468-66-6364)  
On Jan 31, 2001 this sequence version replaced gi:10174886.

COMMENT  
FEATURES

source

```
1. .303249
/organism="Bacillus halodurans"
/db_xref="taxon:86665"
/note="alkaliphile"
104. .1102
/gene="BH2267"
104. .1102
/gene="BH2267"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein"
/protein_id="BAB05986.1"
/db_xref="GI:10174887"
/translation="MKNHLHTIMEDWKLSTALMKKGGDIPFTASLGAFANRAERIPNE
HTRFGIAGCKLFTAIACIQLEAGKLSFDPLSDWLDAPFPNVTHLLHGTSGVP
DYFDEITDFEDLWQVPMHLRLKDLPLFOHAPMKRPPGRRHYNAGKIFLLGL
VYESVGTQFQEVYEAUVFORAGHESGYAEFDLPAKTALGYIDLEDGSGNIFLLYL
PVIGSDGAVTAEADMKWLALMRHELLNETYQKLLPHVHCEDDDDYIGYGVWIK
QQDGAISYHWGDPGVGFHSAFYPTNSGIYVVCANSGSDYVMAAIEALFSEA"
/gene="BH2268"
complement(1146. .1469)
/gene="BH2268"
complement(1146. .1469)
/note="BH2268"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB05987.1"
/db_xref="GI:10174888"
/translation="MKVITLGGSTKFKRQFEQASAYLFLKSHVLSVGFQSESGIEI
TEQVQMFENHRRKIEMSEIFVLDVGGIIGRSTRKEIEFATSKGMPKIRYSESLA
STIFS"
/gene="BH2269"
complement(1636. .2373)
/gene="BH2269"
complement(1636. .2373)
/note="BH2269"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB05988.1"
/db_xref="GI:10174889"
/translation="MVKVGLIVGVFALSIGLMACNGOEVEEKNIDGASGEIPSI
EEIIDINLPDPVATVNGESIDKETVYVLEQOMTMAQYGISLEGEGKOMISLIEE
QTIQELINQLQQAATEKEIYSEIEDEELAFITVAQFSSEALMEALBEGSSMDE
LREIEHYVROQKYVEBETAINVTEETIQAREYEBEKQYTTDEBELPTFEELQSNIBQR
LIAKEQGLVLEKURDGEDITVHI"
/gene="BH2270"
complement(2582. .3367)
/gene="BH2270"
complement(2582. .3367)
/note="BH2270"
/codon_start=1
/transl_table=11
/product="ABC transporter (ATP-binding protein)"
/protein_id="BAB05989.1"
/db_xref="GI:10174890"
/translation="MADILQVEQVTLRKAKKPILOQISWQVEQGEHWAIIVLNGSGKT
SLAKITGVETGEHGVNGLNGVGVHPVIOEVKSGINVSMLDREHRTKGDQTVLEI
VSLGLGTVGVYPLGEADVAKAEAIETFGTLHEDRFLTLSCGERKXAFIARAW
AKPELFDPEPTDGLDIAREQLQLQLESVATTKGAPFLYLVTHYPIEIPFTIHWLM
LKQHVIAQGRDEYLTIKEMITAFAGCSLDVIAQSGRYWVKPM"
/gene="BH2271"
complement(3378. .4310)
/gene="BH2271"
complement(3378. .4310)
/note="BH2271"
/codon_start=1
/transl_table=11
/product="dipeptidase"
```

```
/protein_id="BAB05990.1"
/db_xref="GI:10174891"
/translation="MRFIDAHCDALLKLFADPDAREYTDPAIEANAVRLMEGVRAQC
FAIFVGNPIPDQKQVALQOABLFEKILKKHPOFSIERWDDLILDLADQIGAILT
LEGADPFDDPTKLTLYRLGVSGLTQANLCADGVGPRGGGLTEWGRKAVVSLN
NEHVLTVDSHLSERGFWDYIEAOPVATHSNAKAICKHRNLTDQALKALIEAGGF
IGIVFPHFLTAGTDSDIDDMKHIEHICSLGGQKHVGFSGDFDGPFCFVDSLHAGR
FQWFMDLILKRYDEELVRGEAWDNFLKATPISYV"
4406. .4840
/gene="BH2272"
4406. .4840
/gene="BH2272"
/codon_start=1
/transl_table=11
/product="heat shock protein class I (low molecular
weight)"
/protein_id="BAB05991.1"
/db_xref="GI:10174892"
/translation="MSREESGSKLENVYQDFLRSIDDFSKSQHLHONGMFAPSIP
IRQYKGFQFIIEALFPGVSKNQIQDVPQVYRIRIHEEVLEMRDEKGIYSKSI
SOARERVIPFPVQEGDIKARLNLGIIAIPNHRKRIINTE"
4864. .5193
/gene="BH2273"
4864. .5193
/gene="BH2273"
/note="BH2273"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB05992.1"
/db_xref="GI:10174893"
/translation="MYNLHPROTWSQAYYQFPQHSWPNHPQPTSSGWMQYVSTG
PSPFPFVFNYSQKPKASEPKAAFTNQDGTFLINKTFQVQDVQVKTVMQVPLVKS
VGSIRIK"
complement(5202. .5510)
/gene="BH2274"
complement(5202. .5510)
/note="BH2274"
unknown"
```

Query Match 41.2%; Score 564; DB 1; Length 303249;

Best Local Similarity 64.8%; Pred. No. 8.8e-86;

Matches 837; Conservative 0; Mismatches 455; Indels 0; Gaps 0;

```
QY 1 ATGCGATTGTGAGCCTTATCAGAACCGCTGCAGCGACGATGCAAAAATCGGTGTAAG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 210434 ATGGCGTTTGAAGATTAGCCGAGCGGCTGCAAGATACGTTGAAGAAATTCGCGA 210375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GGTAACTTACTGAAGCTGATATAAAGATAATGATGCGTGAAGTAAGATTAGCGTTATTT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 210374 GGGAAAGTAAGTGACAGACGTAAAGAAATGATGCGTGAAGTCTGCTCTCCTA 210315
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GAGGCTGACGTAACTTTAAAGTGTAAAGAAATTTATTTAAACAGTATCAGAACGSCA 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 210314 GAAGCGGATGTAAACTTTAAAGTGTGAAGCAATTCATCGCAATGTGAAGGAAAAGCC 210255
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 TTAGTTCCGATGTAATGCAATCATTACACCGGCAACAGTTATTAAATAGTTCAA 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 210254 CTCGCTGAAGAAAGTCATGAAGAAAGTTGACACCTGGACACAGGTCTCAAGTCTCAAT 210195
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GATGAATTAACGAAGTGTGGTGGGAGAAAATACATCGATTATATGTCAAATAAACCA 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 210194 GAAGAGCTGACCGCGCTGATGGAGCGCAGCAAGCAAAATTCGGTTCGGAGACGCG 210135
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 CCTACTGTTGTTATGATGGTGGTTTACAGAGTGTGTTAAACACAACTCAGGTAAA 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 210134 CCAACGGCTGCTGATGCTGCTCTTCAAGGGCTGGGAAACGACGACGCGCGAA 210075
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 TTAGCATTTATGATGCGTAAAAAATACACAAAAAACCCTATGTTAGTTCGACAGATTT 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 210074 CTTGCCAACCAATTTTCGGGAAAAAGCACAAATCGGAACCGATTCTCGCTCGCATATT 210015
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 421 TATGTCGACGAGCGTAATAATCAATTAACAACACTAGGGAACAATTAATTCCTGTA 480
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210014 TATGTCGCGCTCGATTAAACAGCTTGAGACATAGGTAACAATAAACATGCTGTG 209955

QY 481 TACAGTGAAGAGATCAACTGAAGCAACCAACAATAATTTACTAATGCAATTAACACGCT 540
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209954 TTTAGCCTTGGTCAATGATAGCGGCGCTTTGAAATTCGCAACAGCGGATCGAGCGGCG 209895

QY 541 AAAGAAGAACATTTAGACATTTGTAATCAATGATACACAGCGTCGATACACATGATGAA 600
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209894 AAAGAAGAGCATCACGACTATGTGCTCATCGATACACAGCTGGAGCTTCCACATCGATGAA 209835

QY 601 GCATTGATGAACGAATTAAGAAGATTAAGAATAATTCCTTAACCCAAACAATTAATGTTA 660
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209834 GAGCTGATGGATGAATCAACAAGTAAGGAAGTGGCCAGCAGATGAATCTTGCTC 209775

QY 661 GTTGTGCAATCAATGACGGGTCAAGATGCTGCAATGCTGCAGAACCTTTTGAGGATCAA 720
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209774 GTCTGTGATGGAATGACAGGCGGATGCGGTTAATGTGCGGAAGCTTCCATGAGCAG 209715

QY 721 CTGTGATCACAGGTGTTTACCTTAACATAAATAGATGATGATACAGGTGGTGGTCACT 780
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209714 CTCGATGTTTCGCGCTGCTTTTGACGAAGCTGGATGGGATCTCGAGCGGAGCTGCC 209655

QY 781 TTATCTATTCTGTCGTGACACAAACCAATTAATTTCTGTTATGATGAGTGAAGAATTA 840
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209654 CTCCTGGTTAAGCGGTTAGAACACGCGGATCAATTTCCGCGGATGGGTGAAGAATTT 209595

QY 841 GATGTTTAGAGCTATTTCATCTCTGAAGTATGTCATCACTGATTTAGTATGCGGTGAT 900
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209594 GACCAAGCTCGACCGCTTCCACCCAGAGCGAATGGCTTCGCGGATTTGGGATGGGGAT 209535

QY 901 GTGTTAAGTTTAATGTAAGAAAGCGCAACAAGATGTGATCAAGAAAGAAACAAAGATTTA 960
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209534 GTGCTGACACTCATTGAAAAAGCAACAATCCACTGCTGATGAAGAAAGCGGCTGAATTT 209475

QY 961 GAGAAAGATGCGTGAATCATGCTTTACTTATGATGATGATGATGATGATGATGATGATG 1020
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209474 GAGAAAGAAATGCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 209415

QY 1021 GTGAAAAATCTAGAACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209414 GTGCGGACATGGGACCACTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 209355

QY 1081 AAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209354 AAAGGCTTGAAGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 209295

QY 1141 CAGTCAATGACCGCGCTGAAGAAACAAATCCAGACACATGATGATGATGATGATGATGATG 1200
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209294 CGTTCATGACGAGCAAGCAAGAAACAAATCCAGACATGATGATGATGATGATGATGATG 209235

QY 1201 CGTATGCTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209234 CGGATCGGAAAGAAAGCGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATG 209175

QY 1261 AACGATATGAAGAAATGATGAAGAAATTCAC 1292
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209174 GAAGAGTGAAGAAAGATGATGAAGAAATTCAC 209143

RESULT 13
AP003191/C 296750 bp DNA linear BCT 10-JUL-2002
LOCUS Clostridium perfringens str. 13 DNA, complete genome, section 7/10.
DEFINITION AP003191 BA000016
ACCESSION AP003191.2 GI:18146729
VERSION
KEYWORDS
SOURCE Clostridium perfringens str. 13 (strain:13) DNA.
ORGANISM Clostridium perfringens str. 13
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
REFERENCE 1
```

```
AUTHORS Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A.,
Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.
TITLE Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
MEDLINE 21664373
PUBMED 11792842
REFERENCE 2 (bases 1 to 296750)
AUTHORS Shimizu,T.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical
Sciences, University of Tsuba, Department of Microbiology; 1-1-1
Tennodai, Tsukuba, Ibaraki 305-8575, Japan
(E-mail:tsimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354,
Fax:81-298-53-3354)
COMMENT On Jan 14, 2002 this sequence version replaced gi:18145205.
FEATURES
Location/Qualifiers
1..296750
/organism="Clostridium perfringens str. 13"
/strain="13"
/db_xref="taxon:195102"
/note="anaerobic pathogen for gas gangrene"
complement(87..773)
/gene="bioB"
complement(87..773)
/gene="bioB"
/note="228 aa, similar to sp:B10D_METJA DETHIOBIOTIN
SYNTHETASE (EC 6.3.3.3) (DETHIOBIOTIN SYNTHASE) (DTB
SYNTHETASE) (DTBS) from Methanococcus jannaschii (248 aa);
33% identity in 221 aa overlap
CPE1543"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="BAB81249.1"
/db_xref="GI:18145206"
/translation="MEKGVVLTGTDIGKTFISGLIKLKEEGNAGYKAVLSGA
IEKNGLLPDCSEVKKISGLKESYENMVSVILENYPISPLASEVEVSIKMKKD
YKSDKXDFILCEGSGVPCPSFSEKMLMEDIIRFNLQIILVNSGSGTINYY
LVSVRLNLGLVKVGIILNKNSDIHRDNKKIIRKELTGVNNISTVPKIEDIKYDL
NELNEVLXGI"
complement(776..1735)
/gene="bioB"
complement(776..1735)
/gene="bioB"
/note="319 aa, similar to pir:E70384 biotin synthetase (BC
2.8.1.6) from Aquifex aeolicus (332 aa); 43.4% identity in
311 aa overlap
CPE1544"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="BAB81250.1"
/db_xref="GI:18145207"
/translation="MDFILKMDKSLKNILKTRFGLRFLNSMLEIKKANIRKEI
HGSDLDLISNGSGRCDFACQKHYKHTNISEYLLDYKIKKAKVENEDEY
HRFSIVTSRGLVGEPEPVITYSNLKNELKINCASHGINKSLIKKAGVKRY
HNLETSRNYDKTCKTHSYEERVKTIKNAEGLVCSGIIIGLGETILDRIDIAIT
LAELEIKSIPINVLISAIGTKLQNLPINEBEILRTIAYPFINPEAKIRLAGGYLL
KNFGNAPKAGANATITGNLLTTCGNKINDDKRLIENIGMRIF"
complement(1754..2308)
/gene="bioY"
complement(1754..2308)
/gene="bioY"
/note="184 aa, similar to sp:B10Y_BACSH BIOY PROTEIN from
Bacillus sphaericus (215 aa); 43.4% identity in 173 aa
overlap. Putative N-terminal signal sequence and 4
putative transmembrane regions were found by PSORT.
CPE1545"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="BAB81251.1"
```

```
/db_xref="GI:18145208"
/translataion="MKMKTRDLTFLIPFAALTAIGAFIKIPIPVPIVPTFLOYFCAQLGA
ILLAGKAGLAQIIVVGLGIGIVPTGGGSPVVPQTFFGLVIGFVIGVILGKITE
NKKINIKNIFLAGILGILVILFVGSYMYVINFILGAMSLLSAVSIGALCLPSD
LILFIIISLVGSKVCPRLKRFYI"
complement(2694..3785)
/gene="CPE1546"
complement(2694..3785)
/gene="CPE1546"
/note="363 aa, similar to sp:YPUA_BACSU HYPOPHETICAL 31.3
KDA PROTEIN IN LISA-PPIB INTERGENIC REGION (ORF19) from
Bacillus subtilis (290 aa); 31% identity in 271 aa
overlap. Putative N-terminal signal sequence was found by
PSORT"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB81252.1"
/db_xref="GI:18145209"
/translataion="MRVLQGAIVSTFMEFGMATPVFAQSEKTEAVVTLGANLTKSERLQ
NLDAGVTKANEVNIIDVTNODIREQLGIDWSKIPASSQSISSVYVVKDGGINVTY
NLDGFTVGSNALLTSGVNADYKADAPKTYTAALAGILKGFEDASGEELSLPK
KEAREEISITNNLSNAKTDGOTLGDRAVVDIKTDYIKDKPKNDEILGIYNN
VTNNYILLTGGQOQETIKPNSKINDLYNFCAMKSLQAMDKLQQLIKDTGQLEE
SGLEKALNIGIKVNLIDKFLVNFSSASEKVKDGTIDENGNIIVIKTNGNSDESKN
RESIQDPQTSNDNNQNEQGNQTK"
complement(4141..7578)
/gene="spovK"
complement(4141..7578)
/gene="spovK"
/note="1145 aa, similar to gp:AB015670_7 Bacillus sp.
genes for Cbase, Cerase, MBP and 15 ORFs, partial and
complete cds from Bacillus sp (549 aa); 45% identity in
522 aa overlap. Also similar to pf:2516401x stage V
sporulation protein K from Bacillus cereus
CPE1547"
/codon_start=1
/transl_table=11
/product="stage V sporulation protein K"
/protein_id="BAB81253.1"
/db_xref="GI:18145210"
/translataion="MLQKNIFVERKDVNKLITENGFFKIEKKLNEEFIGQKTNKEL
CDYFKLEDSKGLIIVEERDIFKNSVLKIFFEELNYIKFKVNSKIDEIDLAAYKF
NVLNATFLDLYSKLINDSQCLVFNTDKASEDILNVLGISPCMLKDNVYIKNK
FLVAEEIDBIVNKLICNNKEFDIFLYNKGDNPEKSGVCTYIEDNFKRGLSLEK
DKNYKIKLLIKNKVKELFDIILIGVDENPDVEKSGVCTYIEDNFKRGLSLEK
CYLSYKICPLINIIRKEAIEANSRLIYVEDIEICKNDENVYDLSKSNPTLEER
YKLSIIGMELKEFDIKENNFKVKIREKGLKTIQSLNIFAGNAGTCKNNAK
ITFFILALGMKENIFKEVSRADFIGEGISNTARRNTEILIESALGGVLFIDEAISC
TDENDKREIVDALLKGIEDNRNITVILAGYEKDMETFLSNQGLKSRFPNYKFE
DYNSEMYEAVNIAKSGYRIAKAGALJELFTKNQVKGSDIGNARFVRNTVENA
IMDAKYLNVKERAIDLLKDNFNKVSADFLEEKIEIIGLEVEKFLRSYKLI
VAQKRSYGVNTKIBONLNIAPNPGTGTSTARVAEMLNSMLGLLVQGVETDR
SSFYSDIPGTSKTEKEKALGVFLIDEAYTLANDSIGREAITLKLIEDYSE
VIVILAGTEEMENPFDVNIGLRSRFLPTKFDYNPNNELLEMALKEAEGFKLSKN
GSEELKNFVEIENADAGSGNRVRNYVEQLIRNOSIKAESDISVYEWNLINTKD
TKAMNAKYDNHDLKLLNLIIGNEELKDFLQYKIMKIREKKKLGFPWDLNKM
NMVFTSGSGCKTFLVNTYSEMLYSGIIRAKNTVQIDKTEFWAMINSKMGVEDILYK
LNYRFLIDWNTFFGEBERYNEIVSDLIKFTDDNSNRIVIVLGGIRGMKDLITN93
HNGRFLVDFGNYNERELYDIASSTLVYKGFITDDEAFTALENALGDIQKRMELAL
XNGLMIQFDSLVRSQSVRCDEDFDINTINRIISDILKSTFELFKKNTIQGNSF"
complement(7634..9622)
/gene="fdh"
complement(7634..9622)
/gene="fdh"
/note="662 aa, similar to pir:S66017 formate dehydrogenase
homolog YyaE from Bacillus subtilis (667 aa); 32.1%
identity in 667 aa overlap
CPE1548"
/codon_start=1
/transl_table=11
/product="probable formate dehydrogenase"
/protein_id="BAB81254.1"
```

```
/db_xref="GI:18145211"
/translataion="MEVLSHGCTLDCDFCKCFNVYEGSEILKIEGDKHEPHTKGLIC
KKGVAHLNRLNKHDRITPLLNKNWVEEISFEDAIEIMKELEYTKYKYSKSLIYV
SOYSGSVGLKIEDIFFNPGYGVSKATGGFCWAGRAQKYDFGVSNSLEDMINSK
NIFUWGNPANTTIIHTMAILNARKKNGSRIIVDINTOSAKLGDHIVHKIPGVDGAL
AMAKIISIKGLQKDFINKYVLGFQFYKDHLENFDLDYLSDECEIIEIKLKY
PKKNSIYLGVMQKNGGNTIRADALGALTQIGVKGGVYANKVILSRILSDY
PKEKSGVENREFYVSNINEETEPKYSLSVEDSNAPIKIWIANSNMNOLPNLR
LNSIDKPEKYCFOMEDTASKDLFIPCTNLTLESDMVPSSNTPNYLIYNEKLTIE
PREKLDMEYFFRELAKPMANLKGYPSLSKKDLYSLIEPLASYNKDITDYLKNEFT
YIDDVAENKKFKTSGFELASKRALKBCGSLTPYLSPRIKENCFLRNLNHSKSL
SSQYIDVKGKAVLNNMIRKFLSLCIGEKVKLKSGTGETITAITCISLDNGVDYVALM
YVGVWKKHGNPNLETSIGSDMGQIITYNETFIEINI"
10033..11127
/gene="prs"
10033..11127
/gene="prs"
/note="364 aa, similar to pir:A71345 probable
phosphoribosyl pyrophosphate synthetase (prs) from
Treponema pallidum (421 aa); 32.3% identity in 325 aa
overlap
CPE1549"
/codon_start=1
/transl_table=11
/product="probable phosphoribosyl pyrophosphate
synthetase"
/protein_id="BAB81255.1"
/db_xref="GI:18145212"
/translataion="MNEQNHGIIALESCTELGNAIDKLIQEKENCEESFLIKTDEI
RFSNGEKVAPQSVNGRDIYILCDIGNSYCTYKMGFENHKGDPDEHQDKRIVSAI
RKGARITVIMPLLYESROHRKRGESLDICALQLEBLRGVDEVITFDVHDHPNQNA
LFGKRRHSRIVNGKNPIVKHYEMGRVDGKDLIVDDMIASGESVLIDIAKELGRNA
RVNVVATFSFTEGLEKFNKFNDDGIKSVYSTNLTITYPEELNAAPWFKAVDLSDFI
ARIINRLYDKSVASVMDATNLIORLIDEX"
complement(11177..12238)
/gene="CPE1550"
complement(11177..12238)
/gene="CPE1550"
/note="353 aa, no significant homology"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB81256.1"
Query Match 38.28; Score 522.8; DB 1; Length 296750;
Best Local Similarity 64.38; Pred. No. 7.1e-79;
Matches 836; Conservative 0; Mismatches 432; Indels 12; Gaps 3;
QY 1 ATGCGATTTGAAGCGCTTATCAGACGCGCTGCAAGCGACGATGCACAAAATCCGCTGTAAG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185046 ATGGCTTTTGATGATTAGCGCTCTAAATTTACAGACACGCTTAAAAAATTAAGGCTAAA 184987
QY 61 GGTAACCTTACTGAACGCTATATAAGATAATGATCGCTGAAGTAAGATTAGCGTTATTT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184986 GGTAAAGCTAACGAAAAAATATAAAAGCCATGAGAGAGTAAGAGCTTGGCGTTTFA 184927
QY 121 GAGCGTACGCTAAACCTTTAAAGTGGTAAAGAAATTTATTAACACATGATCAGAACGCGCA 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184926 GAGCGAGATGTTAACTTCAAGGTGGTAAAAAGCTTTATATCAATGTTAAGATAAATCT 184867
QY 181 TTAGGTTCCGATGTAATGCAATCATATTAACACAGCGCAACAAGTTATTAAAAATAGTCAA 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184866 GTTGGTGAGGAAGTATTAATAAGTTTAACTCCAGGTCACACAGCTTATTAATAATAGTAAAT 184807
QY 241 GATGAATTAACCAAGTTGATGGGTGGAGAAAAATACATCGATTATATATGTCATAATAACCA 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184806 GAGCAATTAACCAACTCTTATGGGAGAGACTGAGAGTAATAATAATACTCAGACAATGGA 184747
QY 301 CCYACTGTGTTATGATGTTGTTTACAAAGTGCCTGTTAAAAACAACAACACTGCAGCTGAAA 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184746 CCACACAGTGTTCATGTTAGTAGTGTCTTCAAGGTGCGAGGTGAAAAACAACACTATGGCTGAAAA 184687
```



SOURCE Streptococcus pyogenes MGAS8232.  
 ORGANISM Streptococcus pyogenes MGAS8232  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 10173)  
 AUTHORS Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,  
 Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklets,S.M.,  
 Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,  
 Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.  
 TITLE Genome sequence and comparative microarray analysis of serotype M18  
 group A Streptococcus strains associated with acute rheumatic fever  
 outbreaks  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)  
 MEDLINE 21927593  
 PUBMED 11917108  
 REFERENCE 2 (bases 1 to 10173)  
 AUTHORS Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,  
 Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklets,S.M.,  
 Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,  
 Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-JAN-2002) Laboratory of Human Bacterial  
 Pathogenesis/Rocky Mountain Laboratories/NIAD/NIH, 903 S. 4th St.,  
 Hamilton, MT 59840, USA  
 FEATURES  
 source Location/Qualifiers  
 1. .10173  
 /organism="Streptococcus pyogenes MGAS8232"  
 /strain="MGAS8232"  
 /db\_xref="taxon:186103"  
 complement(380. .1450)  
 /gene="spyM18\_1147"  
 complement(380. .1450)  
 /gene="spyM18\_1147"  
 /note="best blastp match gb|AAK34058.1| (AE006560)  
 putative integrase/recombinase [Streptococcus pyogenes M1  
 GAS]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative integrase/recombinase"  
 /protein\_id="AAU97765.1"  
 /db\_xref="GI:19748306"  
 /translation="MRRELLLEKITYKAIMPWVLVDYQSKLAIPYFTTLTYEYKE  
 YKRFDWMADATQAPKATDIDLSTLEHFTKDLFAFVLYRERPSLNTYSTKEGSL  
 QKRTNTLSALSSLYLTAEVNDQGEYFIRVNMVKYTKKXETLASRAENIKQF  
 LFGEDTALFDYDYKEYEQKLSNRAKSFKNKREDLAIALLASQVRLSEAVND  
 LKDLNVMITTEYRKGGRDSNVAGFAKGYLESIAVRRYKAEQDLAPFLTYE  
 RQVNRDASSIEKMGVYSEDFKIRVTPHKLRHTLATRLYDATKQVLYNSVHQLGHS  
 TQVTDLYTHINDEQNALDNL"  
 1913. .2611  
 /gene="spyM18\_1150"  
 1913. .2611  
 /gene="spyM18\_1150"  
 /note="best blastp match gb|AAK34059.1| (AE006560)  
 putative repressor protein [Streptococcus pyogenes M1  
 GAS]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative repressor protein"  
 /protein\_id="AAU97766.1"  
 /db\_xref="GI:19748307"  
 /translation="MFSGHLTKATLKSIGTQSELGRLLHVNMKMTISNWEKGRNPNE  
 KHNLAHLFNVSDCFNRYLLPPNQLTISNKEKVIYGERLNLHQIDKSKSLI  
 QGDSQLYAYRVYLSAGTGYSGDFGNFVDYQLEDFASWVFGDSMEYTLTNG  
 EVVLIKNSRFDGCAIYAVENDQGQYIKKVFREDEGLRLVSLNKKYSKDFAPYSESPR  
 IIGKLIANFRPLET"  
 complement(2746. .4308)  
 /gene="fff"  
 /note="spyM18\_1151"  
 complement(2746. .4308)  
 /gene="fff"  
 /note="best blastp match gb|AAK34060.1| (AE006560)  
 putative signal recognition particle -inhibited division  
 protein [Streptococcus pyogenes M1 GAS]"

```
/codon_start=1
/transl_table=11
/product="putative signal recognition particle -inhibited
division protein"
/protein_id="AAL97767.1"
/db_xref="GI:19748308"
/translation="MAFESLQRLODFVKHIRGKKLSQSDVQVETKEIRIALLEADV
ALPVSFTIKRVRERAIGHEIIDIDPTKOILKIVNEELTILGETAEIDKSPKPT
IIMVGLQAGATTAGLANKLIIKEEDARPLAADIYRPAADQLKTDLQGOINPV
FMGTGSDAVIVRGLQARENENDVLDYAGRLQIDDKIMEELRDVLAQAQNEPI
LIVVDSIMQEAANYAVEFNHOLSITGVLTIKDGTGGGALSVELTICKPKIKFTGT
GERITDIETHPDRMSSRIILGDLIIHLIKASQEDDEKKLSLEAKRMENITFDNFD
LEQDQVQNMGGDELLKIPGMAGNPAIANIKVDENQIARRAIVSSMTPABRENPD
KNPAGNPMSSLEGMGCGMPDLSLGGMDMSQLFGGFGFKIGIQFAMQAMKRRQ
ANKLKKAKKRR"
complement(4323..4664)
/gene="yImx"
/note="spyM18_1152"
complement(4323..4664)
/gene="yImx"
/note="best blastp match gb|AAK34061.1| (AE006560)
putative DNA-binding protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/protein_id="AAL97768.1"
/db_xref="GI:19748309"
/translation="WNIMEIKTRNALFEFYAALLTDKQNNYIELYYADDVSLAEI
EAGFVSQVADNIDKTEKILETYEMKHMVSDYVVRSEIFDDMIAHYPHDEYLOEK
ISLTSIDNRE"
complement(4753..5451)
/gene="spyM18_1153"
complement(4753..5451)
/gene="spyM18_1153"
/note="best blastp match gb|AAK34062.1| (AE006560)
putative transcription regulator, GntR family
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative transcription regulator"
/protein_id="AAL97769.1"
/db_xref="GI:19748310"
/translation="MLPAYIKIHDAIKKIDIDGIWPIGSRLPSPERHLAEFTVSRMTL
ROATLLVEGILERRIGSGTVVASHRVQEKMRGTTFTFELLRSQGRQSPSKLSYQK
QLASDTEVKELNDLTDLIRMERIRYADSVPLVYELASIEPEKIKTVKRADITEHP
HSLIANGYIIGSKQTIYAKLASERVASYLEVAKHAILALTQVSYFDGKPEYVRS
QYIGDRFEFYLENN"
complement(5534..6967)
/gene="spyM18_1154"
complement(5534..6967)
/gene="spyM18_1154"
/note="best blastp match gb|AAK34063.1| (AE006560)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL97770.1"
/db_xref="GI:19748311"
/translation="MLRKRILGLFRVELIFGLLLSLVSVYLAWNSPATLHNLIA
TVGIVERSKQPHRYHIGQAVQVQSGPYEQWGTINKQVEDIAENRVSYHYVTFP
IGKVTYSLPELKKRPDKREFKGDIVKLSLTKEPKHVKYQGLATIKQVKKRYDPS
LGQYDILNKLNDLIDGISEQDFVKPYIRFNKNGSPQNNRLIRKAFAYAKAHPNS
VISPGQGHISLPSQDYFELPSDTAIIHQHOTFELHGMKLMFPGPTGPKAEQVR
NLVLTGVHKANDLKKGDHFMIMADHGTDEHVIYDNKFWHKNSHIFDLGSLQNSLF
EKQNFIGYAPELVQQLLSKAQGHDFSEYIOFDAHVHFAWDGGLSLNIAPNFAF
NQRHCHNITYSQNOLFYIDPTGCLRAYSGSIGHSKSGVIRVLNNVFTSSIVTK
AKLTSWFMEPIHFPPNSPVIIVAGNIIN"
7134..8696
/gene="guaA"
/note="spyM18_1155"
7134..8696
```

```
/gene="guaA"
/note="best blastp match gb|AAK34064.1| (AE006560)
putative GMP synthase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative GMP synthase"
/protein_id="AAL97771.1"
/db_xref="GI:19748312"
/translation="MTEISILNDVQKTIIVLDYGSVYNOLIARRIRERGVSSELSKSHI
TAQELREINPIGIVLSSGPNSVYADNAPGIDPEIFELGIPILGICYGMQLITHLGGK
VPYAGAGNREGGQSTLHLRETSKLPSTPEQLVLMHSGDAVTEIPEGHVLGSDND
CPVAATENKMLYIGIOPHVRHSYVDNLIKNFATISCGARGDWSNDFIMETIAK
IRTVGDRLVLGLSGVDSSVGLVQLKAIQGLQCTCFVDRGLLRKDEQDQWGMGLG
GKFLNLIIVDASKRFLDLLADVEDPEKKRIIGNEFVYVDFDASKLKGVDFLAQGT
LYTDITESTETAOIKSHNNVGLPEDMOPFLIBPLNLFKDEVRALGIALGMPBEI
VWQPGPGLAIRVWGAIETREKLEFVRESDAILREETAKAGLDQVQVFTVNTGVR
SVGMGDRGYTYTIAIRATSIDGMTADPAQLPMDVLKAKISTRIVNEVDHVRIVYD
ITSKPPATVEWE"
complement(8738..9961)
/gene="spym18_1156"
complement(8738..9961)
/gene="spym18_1156"
/note="best blastp match gb|AAK34065.1| (AE006560)
putative antimicrobial resistance factor [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative antimicrobial resistance factor"
/protein_id="AAL97772.1"
/db_xref="GI:19748313"
/translation="MTNLTFYAKIGISEEHSDFVKEHQIQISVLOGSDWAKIRNMQN
ERIGTYKEKQVASLSLLIKLLPLGRSIIYIPRGVMDYLDRIYVAFMTKLDYQKT
KALFTKYDPAILLKOYALGOEERBEKPLAALAKNOEAGVHTGLTMEIADSTQPRF
QANLYQENLEMQFPKHTERLIKDAKORGVETVRSQSELHKFKIVSLTEKKNLSL
RNEAYQKLMYTYGKAYILHAKVNIPOKLDIYRQOILILINODITRTOAQHKKRLKL
EDOKASLERITTEFEFTDQYPEVAVAGILSISYGNVEMLYAGNNNDKKFKYPOYL
LYPVQDAYODGIIIVANWGVGEGSLDDGLTKFKNAPTITEFTGEFNFPLVPSLXHI
ANTMYKIRKQLKNKH"
BASE COUNT 3112 a 1914 c 1748 g 3399 t
ORIGIN
Query Match 36.5%; Score 499.2; DB 1; Length 10173;
Best Local Similarity 61.3%; Pred. No. 1.5e-74;
Matches 804; Conservative 0; Mismatches 508; Indels 0; Gaps 0;
QY 1 ATGGCAATTGAGCGCTTATCAGAACCCCTCGAAGCGAGATGCCAAAATGCGGTAG 60
D 4308 ATGGCTTTTGAAGAGCTTAACCCACCGCTTTGCAAGATGTTTAAACACATTCGCGGTA 4249
QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATTAGCGTTATT 120
D 4248 AAGAAATTATCAGAATCAGATGTTCAAGAAGTGACAAAAGAGATTCGCTAGCTCTTTA 4189
QY 121 GAGCGTACGTAACCTTAAAGTGGTAAAGAATTATTATAAAGACATGATCAGAACGCCGA 180
D 4188 GAAGCGGATGTCGCTTTACCAGCTTGTAAACCTTTTATTAAAGCGCTTACGTGAGCGCT 4129
QY 181 TTAGGTCGCGATGAATGCAATCATTAACACCGGCGCAACAGTTATTAATAAGTTCAA 240
D 4128 ATTGGCCATGAGATTATTGATCTCTTGATCCACAACAAATTCCTTAGATGTTAAT 4069
QY 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCAATAAACA 300
D 4068 GAAGAATTAACGAGTATTTTAGGCTCAGAACTCGAGAGATTGATAAATCTCAAAAT 4009
QY 301 CCTACTGTTTATGATGGTGTGTTTACAGGTGCTGGTAAACACACAACTGCAGGTAAA 360
D 4008 CCAACTATTATCATGATGGTGGCTTTGCAAGGGCTGTTGTAACAACTTTTCAGGAAA 3949
QY 361 TTAGCATTTATGATCGCTCAAAAANTACAAAACCTATGTAGTTCGACGATATT 420
D 3948 CTCGCTAATAAATAATTAAGAAGAAGATGCGGACCATGATGATGCTGCCGATATT 3889
```

```
QY 421 TATCGTCCAGCAGCGATTAATCAATACAAACAGTAGGAAACAAATTTGATTCCTGTA 480
Df 3888 TACCGTCCAGCGCGATGATGATGATTAATAACACATGGGCAACAAATCAATGATCTGTT 3829
QY 481 TACAGTGAAGAGAGATCAAGTAAGACCAACAAATTTGATTAATGATTAATAACATGCT 540
Df 3828 TTTGATATGGGACAGACACCACTCAGCAGTCGATATGTCAGAAAGATTTGGACAGGCT 3769
QY 541 AAGAAGAACAATTAGACTTTGTAATGATGATACAGCAGTGCATGATACACATGCAATGAA 600
Df 3768 CGTGAACCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3709
QY 601 GCATTCATGACGATTAAGAGAGTAAGAGAAATTTGTAACCAACCAACCAATTAATGTTA 660
Df 3708 AACTGATGAGAGACTCGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 3649
QY 661 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Df 3648 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3589
QY 721 CTGATGCTCAGAGTGTACCTTAACAAATGATGATGATGATGATGATGATGATGATGATG 780
Df 3588 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3529
QY 781 TTAATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Df 3528 CTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3469
QY 841 GATGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Df 3468 ACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3409
QY 901 GTGTTAAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Df 3408 TTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3349
QY 961 GAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Df 3348 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3289
QY 1021 GTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Df 3288 GTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3229
QY 1081 AAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Df 3228 CAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3169
QY 1141 CAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Df 3168 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3109
QY 1201 CGTATGCTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Df 3108 CGTATGCTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3049
QY 1261 AAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312
Df 3048 AATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2997
```

## RESULT 15

```
U88582
LOCUS U88582 4152 bp DNA linear BCT 03-APR-2001
DEFINITION Streptococcus mutans Prox (opuABC) gene, partial cds; YlxM (ylyxM)
gene, complete cds; Ffh (ffh) gene, complete cds, alternatively
spliced; SatC (satC) and SatD (satD) genes, complete cds; and SatE
(satE) gene, partial cds.
U88582
VERSION U88582.2 GI:11055962
KEYWORDS Streptococcus mutans.
SOURCE Streptococcus mutans
ORGANISM Streptococcus mutans
```

```
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 4152)
Gutierrez,J.A., Crowley,P.J., Brown,D.P., Hillman,J.D., Youngman,P.
and Bleiweis,A.S.
Insertional mutagenesis and recovery of interrupted genes of
Streptococcus mutans by using transposon Tn917: preliminary
characterization of mutants displaying acid sensitivity and
nutritional requirements
J. Bacteriol. 178 (14), 4166-4175 (1996)
96312359
8763945
2 (bases 1 to 4152)
Gutierrez,J.A., Crowley,P.J., Cvitkovitch,D.G., Brady,L.J.,
Hamilton,I.R., Hillman,J.D. and Bleiweis,A.S.
Streptococcus mutans ffh, a gene encoding a homologue of the 54 kDa
subunit of the signal recognition particle, is involved in
resistance to acid stress
Microbiology 145 (Pt 2), 357-366 (1999)
99173740
10075418
3 (bases 1 to 4152)
Kremer,B.H., van der Kraan,M., Crowley,P.J., Hamilton,I.R.,
Brady,L.J. and Bleiweis,A.S.
Characterization of the sat operon in Streptococcus mutans:
evidence for a role of ffh in acid tolerance
J. Bacteriol. 183 (8), 2543-2552 (2001)
21172871
11274114
4 (bases 1 to 4152)
Gutierrez,J.A. and Bleiweis,A.S.
Direct Submission
Submitted (04-FEB-1997) Oral Biology, University of Florida, P.O.
Box 100424, JHMHC, Gainesville, FL 32610-0424, USA
5 (bases 1 to 4152)
Kremer,B.A.H., Crowley,P.J., Gutierrez,J.A. and Bleiweis,A.S.
Direct Submission
Submitted (31-OCT-2000) Oral Biology, University of Florida, 1600
SW Archer Rd, Gainesville,, FL 32610, USA
Sequence update by submitter
On Oct 31, 2000 this sequence version replaced gi:1850604.
Location/Qualifiers
1. 4152
/organism="Streptococcus mutans"
/strain="JH1005"
/db_xref="taxon:1309"
/clone="pJG301"
<1. 414
/gene="opuABC"
<1. 414
/gene="opuABC"
/function="putative glycinebetaine-binding protein"
/notes="synonym=opuAC; component of an ABC transport system
involved in the osmoregulated uptake of glycinebetaine"
/codon_start=1
/transl_table=11
/product="ProX"
/protein_id="AAB48048.1"
/db_xref="GI:1850605"
translation="MKAANKALKSYNSLDKLVSASTGAMTALDQYRNKKDIYVA
AWSPHWMFAKYKLKLDKSKDFGSIHSITRKGLKKIDIPKANKIIDKFNWTKDM
EAVMLDINGMSPEAKAKWIKHPKKVASWTSNK"
524. .856
/gene="ylyxM"
524. .856
/gene="ylyxM"
/notes="putative DNA-binding protein; unique for
Gran-positive microorganisms"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="ylyxM"
/protein_id="AAB48049.1"
```

```
/db_xref="GI:1850606"
/translation="MEYKTNRNALFEYVAAALFDKXOMVIELYVADVSLAEIAEE
EDVSRQAVDNKIKTEKILEDYKMLHMYSDVVESEIFDALMKKYPNDPFLQNKISI
LPTIDNRD"
969..2519
/gene="ffh"
969..2519
/gene="ffh"
/notes="similar to 54 kDa subunit of signal recognition
particle"
/codon_start=1
/transl_table=11
/product="Ffh"
/protein_id="AAB48050.1"
/db_xref="GI:1850607"
/translation="MAPESLTERLQGVFNLRKCKLSEKDVQVETKEIRLALLADV
ALPVKFEIKRVRKRAVHEVIDLDPSSQIIKIYNEBELTAVLGSETAEIKSSKIPT
IIMVGLGCTGATFAGLANKLVKEENARPLMIADYIRPAADQLKILGOQINPV
FDMTEHSAVEIVSQGAQAKENRNDYVLIDTAGRLQIDEXLMTFELDKALANPEI
LIVDSMIGQRAANVAREFNOOLEVTGVILIKIDGTRGGAALSVROITGPKTKFTGT
GKXIDTIEFHPDRMSREILMGDLJTLIEKASODYDQKSAELAEKMRNSFDPNF
IEGLDVOVMGSMEDILKIPGMANKPALANVEYDEGEIARKRAIVSSMTPERNPDP
LTPSRRIASGSGTIVNPKFIKDFQARKMMQGVGDMNKVYKQMGINPNMP
NMDSALBGMGQGMGPMDSGLSGANMDVSMFGGLKGVGEFAMKOSMKNAKRM
KXAKKEK"
2539..3066
/gene="satc"
2539..3066
/gene="satc"
/notes="unknown function"
/codon_start=1
/transl_table=11
/product="SatC"
/protein_id="RAG28335.1"
/db_xref="GI:11055963"
/translation="MNLNLTYSIKRYNGLPNKIQIIIIALSVSILYILPLLISN
GVNGDLHFHIDRMGLSSIKWSKSPVNFYKGVGVNYPFYRYKGVGVNRL
VSSALSKGSVAIHLPNGIHEIRVTVYAVARVAFAISIIIVIIYAVLYHLEK
HRTTEKIRGLFIY"
3163..3831
/gene="satd"
3163..3831
/gene="satd"
/notes="unknown function"
/codon_start=1
/transl_table=11
/product="Satd"
/protein_id="AAG28336.1"
/db_xref="GI:11055964"
/translation="MYIAIIGDLISSKATTNRPKSOKLNLLNOINKYKELLKSA
FTTTGDEPQALLVPNPQIFQIDELALGPKYQIRFGVSGSSILTEINPEQISGDG
PAYWHARAIDYIHKNDYSNHLAVDLETDSQINAILAACEFIKSKWYTYQYEV
IDGLLQGIYEKFSHKMAEKLDLSPSFNKLKSSGLKIYLRKKVATILLNLAIR
KERR"
3828..4152
/gene="satc"
3828..4152
/gene="satc"
/notes="unknown function"
/codon_start=1
/transl_table=11
/product="Satc"
/protein_id="AAG28337.1"
/db_xref="GI:11055965"
/translation="MFKLVISQFLSGNPFVILLIAHFLADFWQSQKMDLKSNNW
TYLIRHLIIVALPILLISVIPHSLVLSLIFLSHLVLDSDSKLLNSFYKDRSFIKT
AVFLID"
BASE COUNT 1412 a 664 c 781 g 1295 t
ORIGIN
Query Match 36.3%; Score 497.2; DB 1; Length 4152;
Best Local Similarity 61.2%; Pred. No. 4e-74;
Matches 802; Conservative 0; Mismatches 508; Indels 0; Gaps 0;
```

```
QY 1081 AABGGCTAGTATAGCTTATATATGAGTGAAGCAAAATTTGATCATATATTAAGCGATTATC 1140
    || || || || || || || || || || || || || || || || || || || || || ||
Db 2049 CCAGCTCTTGCCAATGTCGAAGTTGATGAAGGTGAATTTGCTAGAAAACGGCCATCGTA 2108
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1141 CAGTCAATGACGCGGCTGAAAGAAACAATCCAGACACATTGAATGATCACGTAAAG 1200
    || || || || || || || || || || || || || || || || || || || || || ||
Db 2109 TCATCATGACACCTGAGACGAGTGAATCTGACCTCCTAACTCCGAGCGCAGACGT 2168
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1201 CGTATTGCTAAAGGCTCTGGTTCGTTTCATTACAAGAGTCAATCGTTTGTATGAACAATTT 1260
    || || || || || || || || || || || || || || || || || || || || || ||
Db 2169 CGTATTGCTCTGTTCTGTTGTAATACCTTTTGTAAATGTCATAAATTCATCAAGATTTT 2228
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1261 AAGCATATGAAGAAATGATCAACAAATTCACCTGGTGGCGGTAAAGGTAA 1310
    || || || || || || || || || || || || || || || || || || || || || ||
Db 2229 AATCAGGCTAGAGAGATGACAGGCTGATGCTCTGGTGATATGAATAA 2278
    || || || || || || || || || || || || || || || || || || || || || ||
```

Search completed: February 24, 2003, 23:49:03  
Job time : 5411 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 20:04:35 ; Search time 260 Seconds

(without alignments)  
11848.976 Million cell updates/sec

Title: US-09-943-108A-1

Perfect score: 1368

Sequence: 1 atggcattgaagcttattc.....gtatgaattacacgttttaa 1368

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDs2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1368	100.0	1368	20	AAx27221 S. aureus ffh codi
2	1356.8	99.2	1368	23	AAx27221 Staphylococcus aur
3	1344	98.0	1368	23	AAx27221 Staphylococcus aur
4	1341	98.0	1365	23	AAx27221 Staphylococcus aur
5	790.4	57.8	792	20	AAx27222 S. aureus ffh codi
6	743.2	54.3	1014	24	AAx27222 Staphylococcus epi
c 7	703.4	51.4	721	23	AAx27222 Staphylococcus aur
8	700.8	51.2	319630	24	ABQ67194 Listeria innocua c
c 9	697.6	51.0	2256	24	ABQ70511 Listeria monocytog

c	10	612.4	44.8	644	18	AAV75400 Staphylococcus aur
	11	568.6	41.6	1416	23	AAx51308 Enterococcus faeca
	12	568.6	41.6	1434	23	AAx51308 Enterococcus faeca
	13	565	41.3	6729	20	AAx12974 Staphylococcus aur
c	14	516.2	37.7	521	23	AAx49408 Staphylococcus aur
c	15	516.2	37.7	521	23	AAx49408 Staphylococcus aur
	16	505.6	37.0	1563	24	AAx67087 Streptococcus poly
	17	496	36.3	1395	23	AAx55790 Streptococcus pneu
	18	496	36.3	1569	20	AAx19484 Streptococcus pneu
	19	496	36.3	1569	24	AAx48379 DNA encoding strep
	20	496	36.3	1572	23	AAx55519 Streptococcus pneu
	21	496	36.3	1605	22	AAx90837 2CFE 28 coding seq
c	22	496	36.3	7577	19	AAV52179 Streptococcus pneu
	23	492.8	36.0	1572	22	AAx90724 CFE 28 coding sequ
	24	491.2	35.9	1563	24	AAx67088 Streptococcus poly
	25	486.4	35.6	1136	24	AAx74556 Bacillus lichenifo
c	26	477.6	34.9	2026	19	AAx296317 S. pneumoniae deri
c	27	477.6	34.9	2026	19	AAx296317 S. pneumoniae deri
c	28	463.2	33.9	468	23	AAx50493 Staphylococcus pneu
c	29	443.4	32.4	1389	23	AAx53231 Staphylococcus pneu
c	30	443.4	32.4	1830121	17	AAx42063 Haemophilus influe
	31	433.8	31.7	533	18	AAx75697 Staphylococcus aur
	32	418.2	30.6	640681	24	AAx92787 Buchnera sp. genom
	33	404	29.5	92407	22	AAx28549 Genomic fragment #
	34	370.6	27.1	1362	22	AAx84550 E. coli growth and
	35	370.6	27.1	1362	23	AAx52529 E. coli DNA for ce
	36	360	26.3	1506	23	AAx56354 Salmonella typhi D
	37	341.4	25.0	8367	21	AAx81483 N. meningitidis pa
c	38	341.4	25.0	349980	21	AAx21544 Neisseria meningit
c	39	341.4	25.0	349980	21	AAx21544 Neisseria meningit
c	40	328.4	24.0	330	23	AAx49670 Staphylococcus pneu
c	41	280	20.5	1374	23	AAx54213 Pseudomonas aerugi
c	42	279.4	20.4	910715	20	AAx20248 Borrelia burgdorfe
c	43	268	19.6	4804	20	AAx65141 Ehrlichia sp. E74
c	44	263.6	19.3	292	23	AAx50792 Staphylococcus aur
c	45	260.8	19.1	1230025	20	AAx91990 Nucleotide sequenc

# ALIGNMENTS

RESULT 1	AAx27221
ID	AAx27221 standard; DNA; 1368 BP.
XX	AAx27221;
AC	AAx27221;
XX	28-MAY-1999 (first entry)
DT	S. aureus ffh coding sequence.
DE	Ffh gene; signal recognition particle; SRP; antimicrobial agent; vaccine;
XX	immunological response; gene therapy; infection; otitis media;
KW	conjunctivitis; toxic shock syndrome; septic arthritis; ss.
XX	Staphylococcus aureus.
OS	Staphylococcus aureus.
XX	EP902087-A2.
PN	17-MAR-1999.
XX	24-AUG-1998; 98EP-0306741.
PF	10-SEP-1997; 97US-0927216.
PR	(SMIK ) SMITHKLINE BEECHAM.
XX	Wallis NG;
XX	WPI; 1999-169238/15.
DR	P-PSDB; AAY00910.
XX	New Staphylococcus aureus Signal Recognition Particle (SRP) with



PF 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242579P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
DR WPI; 2001-611495/70.  
XX P-PSDB; AAU37472.  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
PS Claim 27; Seq ID No 8969; 51pp; English.  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1368 BP; 517 A; 187 C; 291 G; 373 T; 0 other;

Query Match 99.2%; Score 1356.8; DB 23; Length 1368;  
Best Local Similarity 99.5%; Pred. No. 2.9e-293;  
Matches 1361; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCATTGAAGGCTTATCAGAACGCTGCAAGCGACGATGCAAAAATCGGTGAAG 60  
DB 1 ATGGCATTGAAGGCTTATCAGAACGCTGCAAGCGACGATGCAAAAATCGGTGAAG 60  
QY 61 GGTAACTTACTGAAGCTGATATAAGATATGATGCGTGAAGTAAAGTATCGCTTATTT 120  
DB 61 GGTAACTTACTGAAGCTGATATAAGATATGATGCGTGAAGTAAAGTATCGCTTACTT 120  
QY 121 GAGGCTCAGCTTAACCTTAAAGTGTAAAGCAATTTATTAACACAGTATCAGACGGCA 180  
DB 121 GAGGCTCAGCTTAACCTTAAAGTGTAAAGCAATTTATTAACACAGTATCAGACGGCA 180  
QY 181 TTAGGTTCCGATGTAATGCAATCATTAAACACCGGCAACAAGTTATTAATAATAGTTCAA 240  
DB 181 TTAGGTTCCGATGTAATGCAATCATTAAACACCGGCAACAAGTTATTAATAATAGTTCAA 240  
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAATACATGCGATTAAATGTCAAAATAACCA 300  
DB 241 GATGAATTAACGAAGTTGATGGTGGAGAAATACATGCGATTAAATGTCAAAATAACCA 300  
QY 301 CCTACTGTTGTTATGATGTTGGTTTACAGAGTGTGGTAAACAAACACTGCAGGTAAA 360  
DB 301 CCTACTGTTGTTATGATGTTGGTTTACAGAGTGTGGTAAACAAACACTGCAGGTAAA 360

QY 361 TTACATTATTGATGCGTAAATAATACACAAAAAACCCTATGTTAGTTGCGAGCATATT 420  
DB 361 TTACATTATTGATGCGTAAATAATACACAAAAAACCCTATGTTAGTTGCGAGCATATT 420  
QY 421 TATCGTCCAGCAGCGATAAATCAATTCAAAACAGTAGTGGGAAAAAATTTGATTTCCGTGA 480  
DB 421 TATCGTCCAGCAGCGATAAATCAATTCAAAACAGTAGTGGGAAAAAATTTGATTTCCGTGA 480  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCAAACAAATTTGTAACCTAAATGCAATTAACACATGCT 540  
DB 481 TACAGTGAAGGAGATCAAGTAAAGCCAAACAAATTTGTAACCTAAATGCAATTAACACATGCT 540  
QY 541 AAAGAAACACATTAGACTTTGTAATCATTTGATACAGACGCTCGATTACACATCGATGA 600  
DB 541 AAAGAAACACATTAGACTTTGTAATCATTTGATACAGACGCTCGATTACACATCGATGA 600  
QY 601 GCATTGATGAACGAATTAAGAAAGAGTAAAGAAATTTCTFAAACCCAAACGAAATTTATGTTA 660  
DB 601 GCATTGATGAACGAATTAAGAAAGAGTAAAGAAATTTCTFAAACCCAAACGAAATTTATGTTA 660  
QY 661 GTTCTGANTTCATGACGGGTCAAGATGCTGCAATGTTGCGAATCTTTTGACATCAA 720  
DB 661 GTTCTGANTTCATGACGGGTCAAGATGCTGCAATGTTGCGAATCTTTTGACATCAA 720  
QY 721 CTTGATGTCACAGGTGTTACCTTAACATAATAGATGGTGATACACGTGGGTGCGAGCT 780  
DB 721 CTTGATGTCACAGGTGTTACCTTAACATAATAGATGGTGATACACGTGGGTGCGAGCT 780  
QY 781 TTATCTATTGCTGCTGCTGACACAAAAACCAATTAATTTGTTGGTATGAGTGAAGTTA 840  
DB 781 TTATCTATTGCTGCTGCTGACACAAAAACCAATTAATTTGTTGGTATGAGTGAAGTTA 840  
QY 841 GATGGTTAGAGCTATTCCATCTCGACAGTATGCGATCAGTATTTAGGTATGGGTGAT 900  
DB 841 GATGGTTAGAGCTATTCCATCTCGACAGTATGCGATCAGTATTTAGGTATGGGTGAT 900  
QY 901 GTGTTAAGTTTAAATGAAGAACGCAACAGATGTGATCAAGAAAAAGCAAGATTTA 960  
DB 901 GTGTTAAGTTTAAATGAAGAACGCAACAGATGTGATCAAGAAAAAGCAAGATTTA 960  
QY 961 GAGAAAAAGATGCGTGAGTCACTGTTTACTTTAGATGATTTTTAGAACAACTTGATCAG 1020  
DB 961 GAGAAAAAGATGCGTGAGTCACTGTTTACTTTAGATGATTTTTAGAACAACTTGATCAG 1020  
QY 1021 GTGAAAAATCTAGGACCAGTGGATGATATTATGAAAATGATTCACAGGTATGAATAAATG 1080  
DB 1021 GTGAAAAATCTAGGACCAGTGGATGATATTATGAAAATGATTCACAGGTATGAATAAATG 1080  
QY 1081 AAAGGCTAGATGAAGCTTAATATGAGTGAAGCAAAATGATCATATTAAAGCGATTATC 1140  
DB 1081 AAAGGCTAGATGAAGCTTAATATGAGTGAAGCAAAATGATCATATTAAAGCGATTATC 1140  
QY 1141 CAGTCAATGACCGCGGTGAAGAAACAAATCCAGACACATTTGAATGTATCAGTAAAAAG 1200  
DB 1141 CAGTCAATGACCGCGGTGAAGAAACAAATCCAGACACATTTGAATGTATCAGTAAAAAG 1200  
QY 1201 CGTATTCCTAAAGGGTCTGGTCTGTTTCAATTCAGAGAGTCAATCGTTTGCATCAACATTT 1260  
DB 1201 CGTATTCCTAAAGGGTCTGGTCTGTTTCAATTCAGAGAGTCAATCGTTTGCATCAACATTT 1260  
QY 1261 AACGATATGAAGAAATGATGAACAAATTCCTGGTGGCGGTAAAGGTAAAAAGGTAAA 1320  
DB 1261 AACGATATGAAGAAATGATGAACAAATTCCTGGTGGCGGTAAAGGTAAAAAGGTAAA 1320  
QY 1321 CGCAATCAATGCAAAATATGTTTAAAGGTATGAATTTACCGTTTTAA 1368  
DB 1321 CGCAATCAATGCAAAATATGTTTAAAGGTATGAATTTACCGTTTTAA 1368  
RESULT 3  
AAS54393  
ID AAS54393 standard; DNA; 1368 BP.  
XX



AC AAS54393;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Staphylococcus aureus DNA for cellular proliferation protein #705.  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
XX  
OS Staphylococcus aureus.  
XX  
PN W0200170955-A2.  
XX  
XX  
PD 27-SEP-2001.  
XX  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
XX  
DR WPI; 2001-611495/70.  
DR P-PSDB; AMU36534.  
XX  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
PS Claim 27; Seq ID No 8030; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pot\_sequences.  
XX  
XX  
SQ Sequence 1368 BP; 520 A; 190 C; 285 G; 373 T; 0 other;

Query Match 98.2%; Score 1344; DB 23; Length 1368;  
Best Local Similarity 98.9%; Pred. No. 2.1e-290;  
Matches 1353; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGGCATTGAAGGCTTATCAGACGCGCTGCAAGCGAGCATGCAAAAATATGCGTGAAG 60  
DB 1 ATGGCATTGAAGGCTTATCAGACGCGCTGCAAGCGAGCATGCAAAAATATGCGTGAAG 60  
QY 61 GGTAACCTTACGAGCTGATATAGATAATGATGCGTGAAGTAAGATAGCGTTATTT 120  
DB 61 GGTAACCTTACGAGCTGATATAGATAATGATGCGTGAAGTAAGATAGCGTTATTT 120  
QY 121 GAGGCTGACGTAACACTTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180

Db 121 GAGGCTGACGTAACACTTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180  
QY 181 TTAGGTTCCGATGTAATGCAATCATTACACACCGGCAACAGTTTATTAAGTAGTCAA 240  
Db 181 TTAGGTTCCGATGTAATGCAATCATTACACACCGGCAACAGTTTATTAAGTAGTCAA 240  
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTATATGTCAAATAAACCA 300  
Db 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATATATCGATTATATGTCAAATAAACCA 300  
QY 301 CCTACTGTTGTTATGATGTTGTTTACAGGTGCTGGTAAACAACTATGTTAGTTCGACAGATAT 420  
Db 301 CCTACTGTTGTTATGATGTTGTTTACAGGTGCTGGTAAACAACTATGTTAGTTCGACAGATAT 420  
QY 421 TATCGTCCAGCAGGATAAATCAATTAACAAACAGTAGGGAACAAATTTGATATTCCTGTA 480  
Db 421 TATCGTCCAGCAGGATAAATCAATTAACAAACAGTAGGGAACAAATTTGATATTCCTGTA 480  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAATTAATCAATTAATAAACATGCT 540  
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAATTAATCAATTAATAAACATGCT 540  
QY 541 AAAGAGAACATTTAGACTTTGTAATCAATTTGATACAGCAGGTGATACACATCGATGAA 600  
Db 541 AAAGAGAACATTTAGACTTTGTAATCAATTTGATACAGCAGGTGATACACATCGATGAA 600  
QY 601 GCATTTGATGACGAATTAAGAAAGTAAAGAAATTTGCTAAACCAACAAATTTATGTTA 660  
Db 601 GCATTTGATGACGAATTAAGAAAGTAAAGAAATTTGCTAAACCAACAAATTTATGTTA 660  
QY 661 GTTGTGATCAATGACGCGGTCAAGATGCTGCAATGTTGCAAGTCTTTTGAACGATCAA 720  
Db 661 GTTGTGATCAATGACGCGGTCAAGATGCTGCAATGTTGCAAGTCTTTTGAACGATCAA 720  
QY 721 CTTGATGTCACAGGTGTTACTTTAACTAAATTTAGATGCTGATACACGTGGTGGTCACT 780  
Db 721 CTTGATGTCACAGGTGTTACTTTAACTAAATTTAGATGCTGATACACGTGGTGGTCACT 780  
QY 781 TTATCTATTTCGTTCCGTCACACAAAACCAATTAATTTGTTGATGAGTGAAGATTA 840  
Db 781 TTATCTATTTCGTTCCGTCACACAAAACCAATTAATTTGTTGATGAGTGAAGATTA 840  
QY 841 GATGTTTGAAGCTATTCCATCCCTGCAACGATGCGATCACTATTTTAGTATGGGTGAT 900  
Db 841 GATGTTTGAAGCTATTCCATCCCTGCAACGATGCGATCACTATTTTAGTATGGGTGAT 900  
QY 901 GTGTTAAGTTTAAATTTGAAAAGCGCAACAGATGCGATCAAGAAAAGCAAAAGATTTA 960  
Db 901 GTGTTAAGTTTAAATTTGAAAAGCGCAACAGATGCGATCAAGAAAAGCAAAAGATTTA 960  
QY 961 GAGAAAAGATGCGTGAAGTCACTGTTTACTTTAGATGATTTTTTTAGAACACTTTGATCAG 1020  
Db 961 GAGAAAAGATGCGTGAAGTCACTGTTTACTTTAGATGATTTTTTTAGAACACTTTGATCAG 1020  
QY 1021 GTGAAAATCTAGGACCACTGGATGATATTTTGAATAATGATTCAGGTATGAATAAATG 1080  
Db 1021 GTGAAAATCTAGGACCACTGGATGATATTTTGAATAATGATTCAGGTATGAATAAATG 1080  
QY 1081 AAAGGCTAGATAGCTTAAATATGAGTGAAGCAAAATTTGATCATATTAAGCGATATTC 1140  
Db 1081 AAAGGCTAGATAGCTTAAATATGAGTGAAGCAAAATTTGATCATATTAAGCGATATTC 1140  
QY 1141 CAGTCAATGACCGCGCTGAAAAGAAACCAATCCAGACACATTTGAATGATTCACGTAAGAG 1200  
Db 1141 CAGTCAATGACCGCGCTGAAAAGAAACCAATCCAGACACATTTGAATGATTCACGTAAGAG 1200  
QY 1201 CGTATTGCTAAAGGCTGTGTCCTTATACAGAAAGTCAATCTGTTGATGAACAAATTT 1260  
Db 1201 CGTATTGCTAAAGGCTGTGTCCTTATACAGAAAGTCAATCTGTTGATGAACAAATTT 1260

Dd 1201 CGTATTGCTAAAGGTCCTGGTCGTTTCATTACAGAAGTCAATCGTTTGATGAACAATTT 1260

QY 1261 AACGATATGAAGAAAATGATGAACAATTCACCTGGTGGCGGTAAAGGTAAAGGTAAA 1320  
 |||||

Dd 1261 AACGATATGAAGAAAATGATGAACAATTCACCTGGTGGCGGTAAAGGTAAAGGTAAA 1320  
 |||||

QY 1321 CGCAATCAAAATGCAAAATATGTTAAAGGATGATCAATTTACCGCTTTTAA 1368  
 |||||

Dd 1321 CGCAATCAAAATGCTTAAAGGATGATCAATTTACCGCTTTTAA 1368  
 |||||

RESULT 4

AA551783

ID AA551783 standard; DNA; 1365 BP.

XX AC AA551783;

XX DT 13-FEB-2002 (first entry)

XX DE Staphylococcus aureus DNA for cellular proliferation protein #200.

XX KW Antisense, ds; prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206948P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlssen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR P-PSDB; AAU33924.

XX PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX PS Claim 27; Seq ID No 4365; 511pp; English.

XX CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence encodes an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 1365 BP; 518 A; 190 C; 285 G; 372 T; 0 other;

Query Match 98.0%; Score 1341; DB 23; Length 1365;

Best Local Similarity 98.9%; Pred. No. 9.8e-290;

Matches 1350; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGCGATTTGAAGGCTTATCAGAACGCTCGAAGCGAGATGCGTAAGTAATTCGCGTGAAG 60  
 |||||

Dd 1 ATGCGATTTGAAGGCTTATCAGAACGCTCGAAGCGAGATGCGTAAGTAATTCGCGTGAAG 60  
 |||||

QY 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATTAGCGTTATT 120  
 |||||

Dd 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATTAGCGTTATT 120  
 |||||

QY 121 GAGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180  
 |||||

Dd 121 GAGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180  
 |||||

QY 181 TTAGGTTCCGATGTAATGCAATCAATTAACACGAGGCAACAGTATTATAAATAGTTCAA 240  
 |||||

Dd 181 TTAGGTTCCGATGTAATGCAATCAATTAACACGAGGCAACAGTATTATAAATAGTTCAA 240  
 |||||

QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTAAATATGTCAAATAACCA 300  
 |||||

Dd 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATATATCGATTAAATATGTCAAATAACCA 300  
 |||||

QY 301 CCTACTGTTGTATGATGGTTGGTTTACAAGTGTCTGGTAAACAAACACTGCAAGTAAA 360  
 |||||

Dd 301 CCTACTGTTGTATGATGGTTGGTTTACAAGTGTCTGGTAAACAAACACTGCAAGTAAA 360  
 |||||

QY 361 TTAGCATTATTGATGCGTAAATAATACACAAAAACCTATGTTAGTTGCAGAGATATT 420  
 |||||

Dd 361 TTAGCATTATTGATGCGTAAATAATACACAAAAACCTATGTTAGTTGCAGAGATATT 420  
 |||||

QY 421 TATCGTCCAGCAGCGATAAATCAATTAACACAGTAGGGAACAAATTTGATTTCTGTGA 480  
 |||||

Dd 421 TATCGTCCAGCAGCGATAAATCAATTAACACAGTAGGGAACAAATTTGATTTCTGTGA 480  
 |||||

QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACCTAATGCAATTAACACATGCT 540  
 |||||

Dd 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACCTAATGCAATTAACACATGCT 540  
 |||||

QY 541 AAAGAAGAACATTTAGACATTTGTAATCATGATACAGAGGTCGATTACACATCGATGAA 600  
 |||||

Dd 541 AAAGAAGAACATTTAGACATTTGTAATCATGATACAGAGGTCGATTACACATCGATGAA 600  
 |||||

QY 601 GCATTGATGAACGAATTAAGAAGAGTAAAGAAATTTGCTAAACCAAGAAATTTATGTTA 660  
 |||||

Dd 601 GCATTGATGAATTAAGAAGAGTAAAGAAATTTGCTAAACCAAGAAATTTATGTTA 660  
 |||||

QY 661 GTTGTGATTCGAATGACGGGTCAAGATGCTGTCAATGTTGCGAGAACTTTTGACGATCAA 720  
 |||||

Dd 661 GTTGTGATTCGAATGACGGGTCAAGATGCTGTCAATGTTGCGAGAACTTTTGACGATCAA 720  
 |||||

QY 721 CTTGATGTCACAGCTTTACCTTAACTAAATAGATGGTGCATACACGTGGTGGCAGCT 780  
 |||||

Dd 721 CTTGATGTCACAGCTTTACCTTAACTAAATAGATGGTGCATACACGTGGTGGCAGCT 780  
 |||||

QY 781 TTATCTATTTCGTTCCGTGACACAAACCAATTAATTTGGTATGAGTGAAGTAAAGTTA 840  
 |||||

Dd 781 TTATCTATTTCGTTCCGTGACACAAACCAATTAATTTGGTATGAGTGAAGTAAAGTTA 840  
 |||||

QY 841 GATGGTTTACAGCTATTCACCTCGAAGTATGTCATCGATCATGATTTTATGAGTATGGGTGAT 900  
 |||||

Dd 841 GATGGTTTACAGCTATTCACCTCGAAGTATGTCATCGATCATGATTTTATGAGTATGGGTGAT 900  
 |||||

QY 901 GTGTAAAGTTTAAATTCAAAAGCGCACACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960  
 |||||

Dd 901 GTGTAAAGTTTAAATTCAAAAGCGCACACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960  
 |||||

QY 961 GAGAAAAAGATGCGTGATCGTTTACTTTAGATGATTTTTTGAACAACCTTTGATCAG 1020  
 |||||

Dd 961 GAGAAAAAGATGCGTGATCGTTTACTTTAGATGATTTTTTGAACAACCTTTGATCAG 1020  
 |||||

QY 1021 GTGAAAATCTAGACACCTGGATGATATTATGAAATGATTCAGGTATGAATAAATG 1080  
 Db 1021 GTGAAAATTTAGACACCTGGATGATATTATGAAATGATTCAGGTATGAATAAATG 1080  
 QY 1081 AAAGGCTAGATAGCTTAATATGATGAGTAAAGCAAAATGATCATATTAAGCGATTATC 1140  
 Db 1081 AAAGGCTAGATAGCTTAATATGATGAGTAAAGCAAAATGATCATATTAAGCGATTATC 1140  
 QY 1141 CAGTCAATGACCGCGCTGAAAGAACAAATCCAGACACATGATGATCAGGTAAAG 1200  
 Db 1141 CAGTCAATGACCGCGCTGAAAGAACAAATCCAGACACATGATGATCAGGTAAAG 1200  
 QY 1201 CGTATTGCTAAAGGCTGCTGCTGCTCATTTACAAAGATCAATCGTTGATGAAACAATTT 1260  
 Db 1201 CGTATTGCTAAAGGCTGCTGCTGCTCATTTACAAAGATCAATCGTTGATGAAACAATTT 1260  
 QY 1261 AACGATATGAAGAAAATGATGAACAAATTCATCTGCTGCGGTAAAGGTAAAGGTAAA 1320  
 Db 1261 AACGATATGAAGAAAATGATGAACAAATTCATCTGCTGCGGTAAAGGTAAAGGTAAA 1320  
 QY 1321 CGCAATCAATGCCAAATATGTTAAAGGTATGAATTTACCGTTT 1365  
 Db 1321 CGCAATCAATGCCAAATATGTTAAAGGTATGAATTTACCGTTT 1365

RESULT 5  
 AAX27222  
 ID AAX27222 standard; DNA; 792 BP.  
 XX AC AAX27222;  
 XX AC AAX27222;  
 DT 28-MAY-1999 (first entry)  
 XX S. aureus ffh coding sequence.  
 DE S. aureus ffh coding sequence.  
 XX Ffh gene; signal recognition particle; SRP; antimicrobial agent; vaccine;  
 KW immunological response; gene therapy; infection; otitis media;  
 KW conjunctivitis; toxic shock syndrome; septic arthritis; ss.  
 XX Staphylococcus aureus.  
 OS Staphylococcus aureus.  
 XX BP02087-A2.  
 XX 17-MAR-1999.  
 XX 24-AUG-1998; 98EP-0306741.  
 XX 10-SEP-1997; 97US-0927216.  
 PR (SMIK ) SMITHKLINE BEECHAM.  
 XX Wallis NG;  
 PI WPI; 1999-169238/15.  
 DR P-PSDB; AAY00911.  
 XX New Staphylococcus aureus Signal Recognition Particle (SRP) with  
 PT protein (ffh) and RNA (ffs) components - the SRP gene and protein  
 PT useful as diagnostic reagents and for prevention and treatment of  
 PT Staphylococci infections which cause otitis media, septic arthritis  
 PT and toxic shock syndrome  
 XX Claim 1; Page 30-31; 35pp; English.  
 PS This sequence encodes the Staphylococcus aureus signal recognition  
 XX particle (SRP) ffh component. Ffh polynucleotides are useful for  
 CC diagnosing a disease related to expression of ffh polypeptides by  
 CC analysing for the presence/amount of ffh protein in a sample due to  
 CC infection of a micro-organism with the gene, or determining the nucleic  
 CC acid sequence encoding ffh. Ffh polypeptides and antagonists are useful  
 CC for treatment of an individual in need (polypeptide) of, or needing to  
 CC inhibit (antagonist) ffh polypeptide levels. Ffh polypeptides and

CC polynucleotides are useful for identifying agonists and antagonists by  
 CC binding and observing the affect of ffh polypeptide activity, which are  
 CC potential anti-microbial agents. Ffh polypeptides and antigenic fragments  
 CC are also useful for inducing an immunological response (T cell/antibody)  
 CC to protect against disease, by direct administration (vaccine), or via a  
 CC vector (gene therapy). Anti-fhh antibodies are useful as antagonists, and  
 CC for protecting against disease. Diseases diagnosed, prevented and treated  
 CC include those caused by infection, especially bacterial infection,  
 CC including otitis media, conjunctivitis, toxic shock syndrome, wound  
 CC infection and septic arthritis. Ffh polypeptides are antimicrobial, and  
 CC are useful for bathing wounds and implants prior to surgical  
 CC implantation.  
 XX

SQ Sequence 792 BP; 296 A; 116 C; 164 G; 216 T; 0 other;

Query Match 57.88; Score 790.4; DB 20; Length 792;  
 Best Local Similarity 99.9%; Pred. No. 6e-167;  
 Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCATTTGAAGGCTTATCAGAAACGCTGCAAGCGAGATGCAAAAATGCGTGTAAAG 60

Db 1 ATGGCATTTGAAGGCTTATCAGAAACGCTGCAAGCGAGATGCAAAAATGCGTGTAAAG 60

QY 61 GGTAAACTTACTGAAGCTGATATAAAGATAATGATGCTGAAGTAAAGTAAAGTAAAGTAAAG 120

Db 61 GGTAAACTTACTGAAGCTGATATAAAGATAATGATGCTGAAGTAAAGTAAAGTAAAGTAAAG 120

QY 121 GAGGCTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180

Db 121 GAGGCTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180

QY 181 TTAGGTTCCGATGTAATGCAATCATTAACACAGGCGCAACAGTTATTAATAAAGTAAAGTAAAG 240

Db 181 TTAGGTTCCGATGTAATGCAATCATTAACACAGGCGCAACAGTTATTAATAAAGTAAAGTAAAG 240

QY 241 GATGAATTAACGAAGTGTGAGTGGTGGAGAAATACATCGATTAATATGTCATAATAACCA 300

Db 241 GATGAATTAACGAAGTGTGAGTGGTGGAGAAATACATCGATTAATATGTCATAATAACCA 300

QY 301 CCTACTGTTTGTATGATGTTGTTTACAAAGTGTGTTTAAACCAACAACTGAGTAAAGTAAAG 360

Db 301 CCTACTGTTTGTATGATGTTGTTTACAAAGTGTGTTTAAACCAACAACTGAGTAAAGTAAAG 360

QY 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACCTATGTTAGTTGACAGATATTT 420

Db 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACCTATGTTAGTTGACAGATATTT 420

QY 421 TATCTCCAGCAGCGATTAATCAATTACAAACAGTAGGGAACAAATGATATTCCTGTA 480

Db 421 TATCTCCAGCAGCGATTAATCAATTACAAACAGTAGGGAACAAATGATATTCCTGTA 480

QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTAACTTAATGCAATTAACCAATGCT 540

Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTAACTTAATGCAATTAACCAATGCT 540

QY 541 AAAGAAGACATTTAGACTTTGTAATCATGATGATGATGATGATGATGATGATGATGATGATGAT 600

Db 541 AAAGAAGACATTTAGACTTTGTAATCATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY 601 GCATTTGATGACGATTAAGAAAGTAAAGAAATTTGCTTAACCAACAAAGTAAATGTTA 660

Db 601 GCATTTGATGACGATTAAGAAAGTAAAGAAATTTGCTTAACCAACAAAGTAAATGTTA 660

QY 661 GTTGTGATTTCAATGACGGTCAAGATGCTGTCATTTGTCAGAAATCTTTTGGACGATCAA 720

Db 661 GTTGTGATTTCAATGACGGTCAAGATGCTGTCATTTGTCAGAAATCTTTTGGACGATCAA 720

QY 721 CTTGATGTCACAGGTGTTTACCTTAACATAATGATGATGATGATGATGATGATGATGATGATGATGAT 780

Db 721 CTTGATGTCACAGGTGTTTACCTTAACATAATGATGATGATGATGATGATGATGATGATGATGATGAT 780

QY 781 TTATCTATTTCTGT 792

Db 781 TTATCTATTTCTGT 792

QY 792 TTATCTATTTCTGT 792

Db 792 TTATCTATTTCTGT 792

Db 781 TTATCTATTGGT 792

RESULT 6  
ABN92495  
ID ABN92495 standard; DNA; 1014 BP.  
XX  
AC ABN92495;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1958.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy; gene; ds.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
DR P-FSDB; ABP39950.  
XX  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 1958; 267pp; English.  
XX  
CC ABN90538 to ABN9374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 1014 BP; 354 A; 156 C; 221 G; 283 T; 0 other;

Query Match 54.3%; Score 743.2; DB 24; Length 1014;  
Best Local Similarity 84.1%; Pred. No. 2.2e-156;  
Matches 838; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 1 ATGGCATTGAAGCTATACAGAGCCCTGCAAGCGATGCAAAAAATGCGGTGAAG 60  
|||||  
Db 16 ATGGCATTGAAGGATTATCCGATCGCTTACAGCCACGATGCAAAAAATGCGGTAAA 75  
|||||  
QY 61 GGTAACTACTGAAGCTATACAGAGATATGATGCTGAAGTATAGATGCGTTATT 120  
|||||  
Db 76 GGAAGATTAACAGAGCAGAGATTAACAGATGATGCTGAAGTATAGATGCGTTATT 135  
|||||  
QY 121 GAGCGTACGTAACCTTTAAAGTGAAGAAATTTATTAACAGATACAGAGCGCA 180  
|||||  
Db 136 GAAGCCGATGTTAACTTCAAAAGTTTGTAAAGATTTTGTAAAGATTTTTCAGAGAGCG 195  
|||||  
QY 181 TTAGTTCGATGTAATGATCATCAACACAGCGCAACAGTATTAATAATAGTCAA 240  
|||||  
Db 196 CTAGGTTCTGATGTCATCTTAAACCTGCGCAACAGGTTATTAATATCGTCAA 255  
|||||

QY 241 GATGAATTACGAGAGTGGTGGAGAAATACATCGATTAAATATGCAAAATAACCA 300  
|||||  
Db 256 GAAGAAGTTTACTAGTTTAAATGGTGGAGAAATACTTCCATTAAAGATGGCAAAACCA 315  
|||||  
QY 301 CCTACTGTTGTTATGATGGTTTCAAGGTGCTGGTAAACCAACCACTGCAGGTAAA 360  
|||||  
Db 316 CCAACTGTTGTCATGATGGTGGCTTACAGGTGCAGGTAAACCAACCACTGCAGGTAA 375  
|||||  
QY 361 TTACATATTGATGCGTAAATAATACACAAACCACTATGTTAGTTGCAAGAGATT 420  
|||||  
Db 376 TTGCATTATTAAATGTTAAATAATATACAAACCACTTACTTTGCGCAGGAGATT 435  
|||||  
QY 421 TATGTCACACGCGATAATCAATACACACAGTAGGAGAAACAAATGATTCCTGTA 480  
|||||  
Db 436 TATGTCACAGCTCTATTGATCAATTACACACAGTAGTAAACAAATGATTCCTGTTG 495  
|||||  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACCAAAATTTGTAATCAATTAATCAATGCT 540  
|||||  
Db 496 TATAGTGAAGGTGATCAAGTATCACCAACCAAAATTTGTAATCAATGCTTAAACATGCT 555  
|||||  
QY 541 AAAGAAGACATTTAGACTTTGTAATCAATGATACAGAGGTGCTGATACATCGATGAA 600  
|||||  
Db 556 AAAGAAGACATCTAGATTTCGTTATCATTTGATACAGCTGCTGTTTACATTTGATGAA 615  
|||||  
QY 601 GCATTGATGAACGAATTAAGAAGTAAAGAAATTCGTAACCAACCAAGAAATTTATGTTA 660  
|||||  
Db 616 GCATTTATGATGAGCTTCAAGAAGTTAAGAATCTTAACCAACCAAGAAATTTATGCTT 675  
|||||  
QY 661 GTTGTGATTCAATGACGGTCAAGATGCTGCAATGTTGAGAAATCTTTGACGATCAA 720  
|||||  
Db 676 GTTGTGATGCAATGACAGGTCAAGATGCTGTAATGTCGACAAATCTTTGATGAA 735  
|||||  
QY 721 CTTGATGTCACAGGTGTTTACCTTAATCAATGATGAGTGGTGAATGCTGCTGAGCT 780  
|||||  
Db 736 TTAGATTGTTTCAAGTGTAACTGATTAATGATGTTGATGATGATGATGATGATGATG 795  
|||||  
QY 781 TTATCTATTGTTGCTGACACAAACCAAAATTTGTTGGTATGATGATGATGATGATGAT 840  
|||||  
Db 796 CTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 855  
|||||  
QY 841 GATGTTTAGAGTATTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
|||||  
Db 856 GATGTTTAGAGTATTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 915  
|||||  
QY 901 GTGTTAGTTTAAATGAAAGCGCAACAGATGATGATGATGATGATGATGATGATGATGAT 960  
|||||  
Db 916 GTTGTAGCTTCAATGAAAGCGCAACAGATGATGATGATGATGATGATGATGATGATGAT 975  
|||||  
QY 961 GAGAAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996  
|||||  
Db 976 GAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011  
|||||

RESULT 7  
AAS49372/c  
ID AAS49372 standard; DNA; 721 BP.  
XX  
AC AAS49372;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Staphylococcus aureus cellular proliferation inhibitory sequence #596.  
XX  
KW Antisense; ss; prokaryotic cellular proliferation;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Staphylococcus aureus.  
XX  
PN W0200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
XX  
DR WPI; 2001-611495/70.  
XX  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX  
PS Claim 1; Seq ID No 1949; 511pp; English.  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence is an antisense  
CC oligonucleotide of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 721 BP; 191 A; 149 C; 104 G; 277 T; 0 other;

Query Match 51.4%; Score 703.4; DB 23; Length 721;  
Best Local Similarity 99.2%; Pred. No. 1.5e-147;  
Matches 707; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGCATTTGAAGGCTTATCAAGACGCTGCAAGCGACGATGCAAAAAATCGGTGTAAG 60  
DB 713 ATGGCATTTGAAGGCTTATCAAGACGCTTGCAGCGACGATGCAAAAAATCGGTGTAAG 654

QY 61 GSTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTATT 120  
DB 653 GGTAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTATT 594

QY 121 GAGGCTGACGTAACATTTAAAGTGGTAAAAAGATTTATTAAACAGATATCAGAACGCGCA 180  
DB 593 GAGGCTGACGTAACATTTAAAGTGGTAAAAAGATTTATTAAACAGATATCAGAACGCGCA 534

QY 181 TTAGGTTCCGATGTAATCAATCATTAACACGAGGCAACAAGTTATTAAAAATAGTTCAA 240  
DB 533 TTAGGTTCCGATGTAATCAATCATTAACACGAGGCAACAAGTTATTAAAAATAGTTCAA 474

QY 241 GATCAATTAACGAGTTGATGGTGGAGAAATACATCGATTATATGTCATTAATCAACCA 300  
DB 473 GATCAATTAACGAGTTGATGGTGGAGAAATACATCGATTATATGTCATTAATCAACCA 414

QY 301 CCTACTCTGTTATGATGGTGGTTTACAAAGGTGGTAAAAACAACACTGCAGGTAAA 360  
DB 413 CCTACTCTGTTATGATGGTGGTTTACAAAGGTGGTAAAAACAACACTGCAGGTAAA 354

QY 361 TTAGCATTTATGATGCGTAAAAAATACAAACAAAAACCTATGTTAGTTGCGAGCATATT 420

DB 353 TTAGCATTTATGATGCGTAAAAAATACAAACAAAAACCTATGTTAGTTGCGAGCATATT 294  
QY 421 TATCGTCCAGCAGCGATAAATCAATTACAAAACAGTAGGAAACAAATTTGATATTCCTGTA 480  
DB 293 TATCGTCCAGCAGCGATAAATCAATTACAAAACAGTAGGAAACAAATTTGATATTCCTGTA 234  
QY 481 TACAGTGAAGGAGATCAAGTAAAGGCCACACAAATTTGAACATAATGATTAACAAATGCT 540  
DB 233 TACAGTGAAGGAGATCAAGTAAAGGCCACACAAATTTGAACATAATGATTAACAAATGCT 174  
QY 541 AAAGAGACACATTAGACTTTGTAATCATTTGATACAGCAGGTGCGATTACACATCGATGA 600  
DB 173 AAAGAGACACATTAGACTTTGTAATCATTTGATACAGCAGGTGCGATTACACATCGATGA 114  
QY 601 GCATTGATCAACGAATTAAGAGAGTAAAGAAATTTGCTAAACCAACGAAATTTATGTTA 660  
DB 113 GCATTGATCAACGAATTAAGAGAGTAAAGAAATTTGCTAAACCAACGAAATTTATGTTA 54  
QY 661 GTTGTGCAATTCATGACGGGTCAAGATGCTGCAATGTTGCGAATCTTTTGA 713  
DB 53 GTTGTGCAATTCATGACGGGTCAAGATGCTGCAATGTTGCGAATCTTTTGA 1

RESULT 8  
ABQ67194  
ID ABQ67194 standard; DNA; 319630 BP.  
XX  
AC ABQ67194;  
XX  
DT 29-AUG-2002 (first entry)  
XX  
DE Listeria innocua contig DNA sequence #7.  
XX  
KW Antibacterial; Listeria; food contamination; mutational analysis;  
KW infection; ds.  
XX  
OS Listeria innocua.  
XX  
PN WO200228891-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 04-OCT-2001; 2001WO-FR03061.  
XX  
PR 04-OCT-2000; 2000FR-0012697.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Kunst F, Glaser P;  
XX  
DR WPI; 2002-332479/37.  
XX  
PT New genomic sequences from Listeria species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators -  
XX  
PS Claim 5; SEQ ID 7; 180pp; French.  
XX  
CC The present invention relates to nucleic acid sequences  
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes  
CC and primers for identification and/or detection of Listeria (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of Listeria (potential therapeutic agents), also for  
CC treating infections by Listeria, and are useful as immunogens in  
CC anti-Listeria vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 319630 BP; 105207 A; 55428 C; 66726 G; 92263 T; 6 other;

Query Match 51.2%; Score 700.8; DB 24; Length 319630;  
Best Local Similarity 70.7%; Pred. No. 2.5e-146;  
Matches 933; Conservative 0; Mismatches 387; Indels 0; Gaps 0;

QY 1 ATGGCAATTAAGAGCTTATCAGACGCTGCAAGCGACGATCAAAAATCGTGSTTAAG 60  
DB 187924 ATGGCAATTAAGAGCTTATCAGACGCTGCAAGCGACGATCAAAAATCGTGSTTAAG 187983

QY 61 GGTAACTTACTGAAGCTGATTAAGATTAAGTATGCGGTGAAGTAAAGTATGCTTTATTT 120  
DB 187984 GGTAACTTACTGAAGCTGATTAAGATTAAGTATGCGGTGAAGTAAAGTATGCTTTATTT 188043

QY 121 GAGCTGACGTAACTTTTAAAGTGTAAAGAAATTTATTAACACAGTATCAGACGCGCA 180  
DB 188044 GAGCTGACGTAACTTTTAAAGTGTAAAGAAATTTATTAACACAGTATCAGACGCGCA 188103

QY 181 TTAGTCTCCGATGTAATCAATCATTAACACGAGGCAACAGTATTAAATAGTTCAA 240  
DB 188104 TTAGTCTCCGATGTAATCAATCATTAACACGAGGCAACAGTATTAAATAGTTCAA 188163

QY 241 GATGAATTAACGAAGTTGGTGGGAGAAATACATCGATTAATATGTCAAAATAACCA 300  
DB 188164 GATGAATTAACGAAGTTGGTGGGAGAAATACATCGATTAATATGTCAAAATAACCA 188223

QY 301 CCTACTGTTGATGATGTTTACAGGCTGCTGTTAAACACACACTGCGAGTAA 360  
DB 188224 CCTACTGTTGATGATGTTTACAGGCTGCTGTTAAACACACACTGCGAGTAA 188283

QY 361 TTAGCATATTGATGCTAAAAATACACAAAAAOCCTGTAGTTGCGACGAGATTT 420  
DB 188284 TTAGCATATTGATGCTAAAAATACACAAAAAOCCTGTAGTTGCGACGAGATTT 188343

QY 421 TATGCTCCAGCAGTAAATCAATTAACACAGTAGGGAACAAATGATATTCCTGTA 480  
DB 188344 TATGCTCCAGCAGTAAATCAATTAACACAGTAGGGAACAAATGATATTCCTGTA 188403

QY 481 TACAGTGAAGAGATCAAGTAAAGCAACACAAATTTACTAATGCAATTAACATGCT 540  
DB 188404 TACAGTGAAGAGATCAAGTAAAGCAACACAAATTTACTAATGCAATTAACATGCT 188463

QY 541 AAAGAAGACATTTAGACATTTTGTATCATTTGATACAGAGCTCGATTAACATCGATGAA 600  
DB 188464 AAAGAAGACATTTAGACATTTTGTATCATTTGATACAGAGCTCGATTAACATCGATGAA 188523

QY 601 GCATGATGAACGAATTAAGAAATTAAGAAATTTGCTTAACCAACGAAATATGTTA 660  
DB 188524 GCATGATGAACGAATTAAGAAATTAAGAAATTTGCTTAACCAACGAAATATGTTA 188583

QY 661 GTTGTGATCAATGACGGGTCAAGATGCTGCTCAATGTTGAGAAATCTTTTGAAGATCAA 720  
DB 188584 GTTGTGATCAATGACGGGTCAAGATGCTGCTCAATGTTGAGAAATCTTTTGAAGATCAA 188643

QY 721 CTTGATGTCAGAGTGTACCTTTACTTAATTAATAGATGGTGATACAGTGGTGGTACAGCT 780  
DB 188644 CTTGATGTCAGAGTGTACCTTTACTTAATTAATAGATGGTGATACAGTGGTGGTACAGCT 188703

QY 781 TTACTATTCGTTGCTGACACAAACCAATTAATTTGTTGGTATGATGAGTAAAGTTA 840  
DB 188704 TTACTATTCGTTGCTGACACAAACCAATTAATTTGTTGGTATGATGAGTAAAGTTA 188763

QY 841 GATGTTTATGAGTATTCCTCTGAAAGCTATGCAATCATGATTTTATGATGAGTGGTAT 900  
DB 188764 GATGTTTATGAGTATTCCTCTGAAAGCTATGCAATCATGATTTTATGATGAGTGGTAT 188823

QY 901 GTGTTAGTTTAAATGAAGAAGCGCAACAGATGATGATCAAGAAAGCAAAAGATTTA 960  
DB 188824 GTGTTAGTTTAAATGAAGAAGCGCAACAGATGATGATCAAGAAAGCAAAAGATTTA 188883

QY 961 GAGAAAGATGCGTGAATCATGTTTACTTTAGATGATTTTATGATCAATGATGATGATGAT 1020  
DB 188884 GAGAAAGATGCGTGAATCATGTTTACTTTAGATGATTTTATGATCAATGATGATGATGAT 188943

QY 1021 GTGAAAATCTAGACCACTGATGATGATTAATGAAATGATTCAGGATGATGAATAATG 1080  
DB 188944 GTGAAAATCTAGACCACTGATGATGATTAATGAAATGATTCAGGATGATGAATAATG 189003

QY 1081 AAAGGCTAGATGAAGCTTAATGATGAGTGAAGCAAAATGATCATATTAAGCGATTATC 1140  
DB 189004 AAAGGCTAGATGAAGCTTAATGATGAGTGAAGCAAAATGATCATATTAAGCGATTATC 189063

QY 1141 CAGTCAATGACGGCGCTGAAAGAAACAAATCCAGACACATGTAATGATCAAGTAAAG 1200  
DB 189064 CAGTCAATGACGGCGCTGAAAGAAACAAATCCAGACACATGTAATGATCAAGTAAAG 189123

QY 1201 CGTATTCCTAAAGGCTGCTGCTGTTCAATCAAGAAAGTCAATGCTTTGATGAAACAATTT 1260  
DB 189124 CGTATTCCTAAAGGCTGCTGCTGTTCAATCAAGAAAGTCAATGCTTTGATGAAACAATTT 189183

QY 1261 AACGATTAAGAAATGATGAACAAATTCATGCTGGTGGGTAAGGTAAGGTAAGGTAAG 1320  
DB 189184 AACGATTAAGAAATGATGAACAAATTCATGCTGGTGGGTAAGGTAAGGTAAGGTAAG 189243

RESULT 9  
ABQ70511/c  
ID ABQ70511 standard; DNA; 2256 Bp.  
XX  
AC ABQ70511;  
XX  
DT 29-AUG-2002 (first entry)  
XX  
DE Listeria monocytogenes 4b contig DNA sequence #453.  
XX  
KW Antibacterial; Listeria; food contamination; mutational analysis;  
infection; ds.  
XX  
OS Listeria monocytogenes 4b.  
XX  
PN WO200228891-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 04-OCT-2001; 2001WO-FR03061.  
XX  
PR 04-OCT-2000; 2000FR-0012697.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Kunst F, Glaser P;  
XX  
WPI; 2002-332479/37.  
XX  
PT New genomic sequences from Listeria species, useful for detection,  
treatment and prevention of infection, also related polypeptides,  
antibodies and modulators -  
XX  
PS Claim 14; SEQ ID 3324; 180pp; French.  
XX  
CC The present invention relates to nucleic acid sequences  
CC (ABQ7188-ABQ71212) from Listeria sp. The sequences are useful as probes  
CC and primers for identification and/or detection of Listeria (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of Listeria (potential therapeutic agents), also for  
CC treating infections by Listeria, and are useful as immunogens in  
CC anti-Listeria vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2256 Bp; 533 A; 471 C; 420 G; 832 T; 0 other;

Query Match 51.0%; Score 697.6; DB 24; Length 2256;



Db 301 CCGACAGTGATTATGATGACAGG

Db 301 CCGACAGTGATTATGATGACAGGGTTACAAGGGCTGGTAAACAACCTTTTACTGGTAAA 360



QY 361 TTACGATTATGTCCTGTAATAAATAATACACAAAAACCTATGTTAGTTGCGAGCATATTT 420  
Db 361 TTAGAAAACACATATATGAAAGCTGCTCGCCGCTTTTATCGCTGGTGGACGTT 420  
QY 421 TATCGTCCAGCAGATAAATCAATTAACAACAGTAGGGAACAATGATATTCCTGTA 480  
Db 421 TATCGTCCAGCAGATTAATGATGAGTGAAGGTTTATAGGTCACAAATTAGAAGTTCCCGGT 480  
QY 481 TACAGTGAAGGAGATCAAGTAAGCCACAAATTTGAACATTAATTAACACATGCT 540  
Db 481 TTTGATATGGACAGATGCTAATCCAGTGGAAATTTGTCGTCAGGGTTAGCATTAGCA 540  
QY 541 AAAGAGACATTTAGCTTTGTAATCATGATACACAGAGTGCATACACATGATGAA 600  
Db 541 AAAGAAAGAAAATGANTATGCTTAAATGATACGCGCGCGTTTACACATTCAGCAA 600  
QY 601 GCATGATGAACGAATTAAGAGATTAAGAAATTTGCTTAACCAACAGAAATTAAGTTA 660  
Db 601 GCTTTAATGACGAATTAAGAAATTAAGAGATTTGGCTAATCCCAATGAATTTCTGTTA 660  
QY 661 GTTGTGATTCATGACGGGTCAAGATGCTGCAATGTTGCAGAAATCTTTTACGATCAA 720  
Db 661 GTTGTGATGCGATGACGGGCAAGATGCTGCAAGTTGCGAGATGTTTAAATGAACAG 720  
QY 721 CTTGATGTCACAGGTGTACCTTAACATAATGATGGTGATACACGTGGTGGCAGCT 780  
Db 721 CTTGGAATTAAGTGGGTGTTTATTAACAAAATGGACGGGATACCTCGTGGGGGGCTGCG 780  
QY 781 TTATCTATTCGTCGTCGACACAAACCAATTAATTTGTTGATGATGATGAAAGTTA 840  
Db 781 CTGTCGAATTCGGCGATGACGGGGCGCTCCGATTAATTAATTTGTCGGTCTGGTGAATA 840  
QY 841 GATGGTTTAGAGATTTCCATCCTGAACGTATGCGATCATGATTTTAGGTATGGGTGAT 900  
Db 841 ACCGATTTAGAAATTTCCATCCCGATCGTATGTCGAGTCGTATCCTAGGTATGGGGAC 900  
QY 901 GTGTTAAGTTTAAATGAAGGCGCAACAGATGTGGATCAAGAAAAGCAAAAGATTTA 960  
Db 901 ATGTTGACGCTAATTTGAAAAGCGCAACAGATTACGATGAGAAAAGCAGAGAATTT 960  
QY 961 GAGAAAAGATGCGTCAGTCATCGTTTACTTTAGATGATTTTTAGACACATTTGATCAG 1020  
Db 961 GCTCAAAAATGAAGAAAACAGATTTTGAATTTAAGCATTTCAATGACCAATTTGATCAA 1020  
QY 1021 GTGAAAATCTAGGACCATGATGATTAATGAAATGATTCAGATGATCAATGAAATG 1080  
Db 1021 GTTATGGCATGGGCCGATTTGAAGACTTATTAATGATCCCTGGATGATGATGATG 1080  
QY 1081 AAAGGGTAGATGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
Db 1081 CCTGATTTGAAAATGTCAAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
QY 1141 CAGTCAATGACGGCGCTGAAGAAACAAATCCAGACACATTTGAATGATGATGATGATGAT 1200  
Db 1141 CTATCAATGACCGCTGACAGCGTGAATTCCTGATCTAATTAATCCTAGTCGCGCGTCG 1200  
QY 1201 CGTATGCTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
Db 1201 AGAATTCGACCTGGTTCAGGAAATAGTGTGTTGAAGTCAATCGTATGATGATTAACATTT 1260  
QY 1261 AACGATGATGAAGAAATGATGAACAAATTCACGTGGTGGCGGTA 1303  
Db 1261 AAAGAAATCCAAAATAATGATGCAACAAATGTGTCAAAAGGGGATA 1303

RESULT 12

AAS53127

ID AAS53127 standard; DNA; 1434 BP.

XX AAS53127;

AC AAS53127;

XX 13-FEB-2002 (first entry)

XX

XX

XX

DE Enterococcus faecalis DNA for cellular proliferation protein #555.  
XX Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX Enterococcus faecalis.  
OS  
PN WO200170955-A2.  
XX 27-SEP-2001.  
PD  
XX 21-MAR-2001; 2001WO-US09180.  
PF  
XX 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX (ELIT-) ELITRA PHARM INC.  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;  
PI Yamamoto RT, Xu HH;  
XX WPI: 2001-611495/70.  
DR P-FSDB; AAU35268.  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
PT  
XX Claim 27; Seq ID No 6764; 511pp; English.  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1434 BP; 500 A; 230 C; 340 G; 364 T; 0 other;

Query Match 41.6%; Score 568.6; DB 23; Length 1434;  
Best Local Similarity 64.8%; Pred. No. 2.1e-117;  
Matches 844; Conservative 0; Mismatches 459; Indels 0; Gaps 0;

QY 1 ATGCGATTTGAAGGCTTATPACAGAGCGCTGCAACGCGACGATGCAAAAATCGGTGTAAG 60  
Db 1 ATGCGATTTTGAAGGCTTATPACAAACGCGCTACAAACAGCAATGAGTAAAAATCCGTCGTAAG 60  
QY 61 GGTAACTTACTGAAGCTCATATAAAGATATGATGCTGAAGTATGATTCAGCTTATTT 120  
Db 61 GGAAGAAGTTTCCGAGCGCGACGTAATAAGAAATGATGCGAGAATCCGTTTGGCTTTATTA 120  
QY 121 GAGGCTGACGTAACTTTAAAGTGTGTAAGAAATTTATTATAAACAGATATCAGAACGGCA 180  
Db 121 GAAGCGACGCTTAATTTACAGAGTGTGTAAGAAATTTTCAAAAACGCGTCAGAGAACGGCA 180  
QY 181 TTAGGTTCCGATGTAATGCAATTAACACCGACGAGGCAAAAGTTATTAAAAATAGTTCAA 240

Db 181 GTAGGAGTCGAAGTATTAGAAAGCTTATCACCAGCCCAACAATTCGTAATAATGTTGAT 240  
QY 241 GATGAATTAACGAATGATGGTGGGAGAAATACATCGATTAAATATGTCATAAATACCA 300  
Db 241 GATGAATTAACGAATGATGGTGGGAGAAATACATCGATTAAATATGTCATAAATACCA 300  
QY 301 COTACTGTGTATGATGTTGTTTACAAGGTCGTGTAACAAACAACTGCAAGGTA 360  
Db 301 CGACAGTCATTATGATGACAGGTTTACAGGGGCTGTGTAAACAACTTACTGGTAA 360  
QY 361 TTAGCAATTAATGATGCGGTAAATAACACAAAAACCTATGTTAGTTCACAGATATT 420  
Db 361 TTAGCAAAACACTTAATGAAACCTGAAACAGCTCGTCCGCTTTTATCGTGGTGA 420  
QY 421 TATCGTCCAGCAGGATTAATCAATTAACAACAGTACAGGAAACAAATTCATATTC 480  
Db 421 TATCGTCCAGCAGGATTAATCAATTAACAACAGTACAGGAAACAAATTCATATTC 480  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATTAATCAATTAACATG 540  
Db 481 TTGTATGGAACAGATGCTTAATCCAGTGGAAATTTGTCGTCAGGGTAGCATAGCA 540  
QY 541 AAAGAAACATTTAGACTTTTAATCAATTTGATACACAGGTCGATTCACATCGATGAA 600  
Db 541 AAAGAAAGAAAAATGATGATGTTTAAATGATACGCGCGCGTTCACATTCAGCAA 600  
QY 601 GCATGTATGACAGCAATTAAGAAAGTAAAGAAATTCCTTAACCAACCAAAATTA 660  
Db 601 GCTTTATGGAAGCAATTAAGAAAGTAAAGAAATTCCTTAACCAACCAAAATTA 660  
QY 661 GTGTGCAATTCATGAGCGGTCAAGATGCTCAATGTTGCAAGATCTTTTCAACGATCA 720  
Db 661 GTGTGATGCGATGACGCGGCAAGATGCTCAACGTTCCAGATAGTTTAAAGACAG 720  
QY 721 CTGTATGACAGGTGTTACCTTAATAATAGATGATGATACAGTGGTGGTGCAGCT 780  
Db 721 CTGTGAATTAATGCGGTTTATTAACCAAAATGAGCGCGATACTCGTGGGGGCGTGC 780  
QY 781 TTATCTATTCGTGCGTGACACAAACCAATTAATTTGTTGGTATGATGAAGAATTA 840  
Db 781 CTGTCAATTCGGGCAATGAGCGGCGCTCGATTAATTTGCTGGTTCTCGGTAATAATTA 840  
QY 841 GATGTTTATGAGCTATTCCATCTGCAAGTATGCAATGATGATGATGATGATGATGAT 900  
Db 841 ACCGATTTAGAAATTTCCATCCGATCGTATGTCGAGTCTATCTATGATGATGATGAT 900  
QY 901 GTGTTAGTTAATGAAAGCGCAACAGATGATGATGATGATGATGATGATGATGATGAT 960  
Db 901 ATGTTGAGCTTAATGAAAGCGCAACAGATGATGATGATGATGATGATGATGATGAT 960  
QY 961 GAGAAAGAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
Db 961 GCTCAAAAATGAAAGAAACAGTGTGACCTTAACGATTCATGAGCAATGATGATGAT 1020  
QY 1021 GTGAAAAATCTAGACACTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
Db 1021 GTTATGGCATGGGACCGATGAGACATTAATAAATGATCCCTGGAATGATGATGATGAT 1080  
QY 1081 AAAGGGTAGATTAAGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
Db 1081 CTTGTATGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
QY 1141 CAGTCAATGACCGGCTGAAAGAAACATCCAGACATGATGATGATGATGATGATGATGAT 1200  
Db 1141 CTATCATGACCCCTGACAGAGCTGAAATCCTGATCTATTAATCCTTATGTCGCGTGC 1200  
QY 1201 CGTATTGTAAGGCTGCTGCTCAATTAACAAGAGCTCAATCGTTGATGATGATGATGAT 1260  
Db 1201 AGAATTCAGCTGCTGACGAAATAGTGGTGAATCAATCGTATGATGATGATGATGAT 1260  
QY 1261 AACGATATGAAGAAATATGATGAACAAATTCATCGTGGCGGTA 1303

Db 1261 AAAGAATCCAAAAAATGATGCAACAAATGTCCAAAGGGGATA 1303  
RESULT 13  
AA12974  
ID AA12974 standard; DNA; 6729 Bp.  
XX  
AC AA12974;  
XX  
DT 19-MAR-1999 (first entry)  
XX  
Enterococcus faecalis genome contig SEQ ID NO:37.  
XX  
Enterococcus faecalis; contig; detection; Enterococcal infection;  
KW  
vaccine; attenuation; computer readable medium; ds.  
XX  
Enterococcus faecalis.  
XX  
W09850555-A2.  
PN  
12-NOV-1998.  
XX  
04-MAY-1998; 98W0-US09885.  
XX  
14-NOV-1997; 97US-0066009.  
PR  
06-MAY-1997; 97US-0044031.  
PR  
16-MAY-1997; 97US-0046655.  
XX  
(HUMA-) HUMAN GENOME SCI INC.  
XX  
Barash SC, Dillon PJ, Kunsch CA;  
PI  
WPI; 1999-045171/04.  
DR  
XX  
New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
XX  
Claim 1; Page 406-409; 2084pp; English.  
PS  
XX  
A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC AA12938 to AA13919 represent these nucleotide sequences which are  
CC primary nucleotide sequences, also known as contigs. The computer-based  
CC system can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
XX  
SQ Sequence 6729 Bp; 2131 A; 1181 C; 1357 G; 2053 T; 7 other;  
Query Match 41.3%; Score 565; DB 20; Length 6729;  
Best Local Similarity 64.5%; Pred. No. 2e-116;  
Matches 841; Conservative 1; Mismatches 461; Indels 0; Gaps 0;  
QY 1 ATGGCATTTGAAGCTTATCAGAAGCGCTGCAAGCGAGATGCAAAAATCGTGGTAAG 60  
Db 3722 ATGGCTTTGAGAGTTTACAAACCGCCTACGAGCAATGATTAATCCGTCGTAAG 3781  
QY 61 GGTAACCTTACTGAAGCATATATAAGATGATGCGTGAAGTAAAGATTAGCGTTATTT 120  
Db 3782 GGAAAGTTCCGAAACCGACGTAAGAAATGATGCGAGAAATCCGTTGCTTATTA 3841  
QY 121 GAGGCTGACGTAACTTTAAAGTGTGTAAGAAATTTATTTAAACAGTATCAACAGCGCA 180  
Db 3842 GAAGCCGACGTTAATTTTACAAGTGTCAAGATTTTCAAAAACGTTGTCAGAGAACGGCA 3901



Db 521 AGGGTAAACTTACTGAAGCTGATATAAAGATAATGATCGCTGAAGTAAGATTAGCGTTAC 462  
QY 119 TTGAGGCTGACGTAAGCTTAAAGTGGTAAAGAAATTTATTAACAGATACAGAACCG 178  
Db 461 TTGAGGCTGACGTAAGCTTAAAGTGGTAAAGAAATTTATTAACAGATACAGAACCG 402  
QY 179 CATTAGGTTCCGATGTAATGCAATCAATTAACACAGGCGCAACAGTTATTAATAATAGTTC 238  
Db 401 CATTAGGTTCCGATGTAATGCAATCAATTAACACAGGCGCAACAGTTATTAATAATAGTTC 342  
QY 239 AAGATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCATAATAAC 298  
Db 341 AAGATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCATAATAAC 282  
QY 299 CACCTACTGTTGTTATGATGGTGGTTTACAGGTGCTGTAACAAACAACTGCAGGTA 358  
Db 281 CACCTACTGTTGTTATGATGGTGGTTTACAGGTGCTGTAACAAACAACTGCAGGTA 222  
QY 359 AATTAGCATTTATGATCGCTGTAATTAACAAACAACTGTTAGTTGCGAGGATA 418  
Db 221 AATTAGCATTTATGATCGCTGTAATTAACAAACAACTGTTAGTTGCGAGGATA 162  
QY 419 TTATGCTGCCAGCAGCGATATAATCAATTAACAAACAGTAGGGAACAAATTTGATATTCCTG 478  
Db 161 TTATGCTGCCAGCAGCGATATAATCAATTAACAAACAGTAGGGAACAAATTTGATATTCCTG 102  
QY 479 TATACAGTGAAGGAGATCAAGTAAGGCCACACAAATTTGAACTAATGATTAACCAATG 538  
Db 101 TATACAGTGAAGGAGATCAAGTAAGGCCACACAAATTTGAACTAATGATTAACCAATG 42  
QY 539 CTAAAGAGAACATTTAGACTTTGTAATCATTTGATACAGCA 579  
Db 41 CTAAAGAGAACATTTAGACTTTGTAATCATTTGATACAGCA 1

RESULT 15  
AAS49436/c  
ID AAS49436 standard; DNA; 521 BP.  
XX  
AC AAS49436;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Staphylococcus aureus cellular proliferation inhibitory sequence #660.  
XX  
KW Antisense; ss; prokaryotic cellular proliferation;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
DR WPI; 2001-611495/70.  
XX  
PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
PS Claim 1; Seq ID No 2013; 51lpp; English.  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence is an antisense  
CC oligonucleotide of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 521 BP; 139 A; 102 C; 74 G; 206 T; 0 other;

Query Match 37.7%; Score 516.2; DB 23; Length 521;  
Best Local Similarity 99.4%; Pred. No. 8.2e-106;  
Matches 518; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 59 AGGGTAAACTTACTGAAGCTGATATAAAGATAATGATCGCTGAAGTAAGATTAGCGTTAT 118  
Db 521 AGGGTAAACTTACTGAAGCTGATATAAAGATAATGATCGCTGAAGTAAGATTAGCGTTAT 462  
QY 119 TTGAGGCTGACGTAAGCTTAAAGTGGTAAAGAAATTTATTAACAGATACAGAACCG 178  
Db 461 TTGAGGCTGACGTAAGCTTAAAGTGGTAAAGAAATTTATTAACAGATACAGAACCG 402  
QY 179 CATTAGGTTCCGATGTAATGCAATCAATTAACACAGGCGCAACAGTTATTAATAATAGTTC 238  
Db 401 CATTAGGTTCCGATGTAATGCAATCAATTAACACAGGCGCAACAGTTATTAATAATAGTTC 342  
QY 239 AAGATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCATAATAAC 298  
Db 341 AAGATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCATAATAAC 282  
QY 299 CACCTACTGTTGTTATGATGGTGGTTTACAGGTGCTGTAACAAACAACTGCAGGTA 358  
Db 281 CACCTACTGTTGTTATGATGGTGGTTTACAGGTGCTGTAACAAACAACTGCAGGTA 222  
QY 359 AATTAGCATTTATGATCGCTGTAATTAACAAACAACTGTTAGTTGCGAGGATA 418  
Db 221 AATTAGCATTTATGATCGCTGTAATTAACAAACAACTGTTAGTTGCGAGGATA 162  
QY 419 TTATGCTGCCAGCAGCGATATAATCAATTAACAAACAGTAGGGAACAAATTTGATATTCCTG 478  
Db 161 TTATGCTGCCAGCAGCGATATAATCAATTAACAAACAGTAGGGAACAAATTTGATATTCCTG 102  
QY 479 TATACAGTGAAGGAGATCAAGTAAGGCCACACAAATTTGAACTAATGATTAACCAATG 538  
Db 101 TATACAGTGAAGGAGATCAAGTAAGGCCACACAAATTTGAACTAATGATTAACCAATG 42  
QY 539 CTAAAGAGAACATTTAGACTTTGTAATCATTTGATACAGCA 579  
Db 41 CTAAAGAGAACATTTAGACTTTGTAATCATTTGATACAGCA 1

Search completed: February 24, 2003, 22:24:30  
Job time : 619 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 21:39:11 ; Search time 85 Seconds  
(without alignments)  
4935.692 Million cell updates/sec

Title: US-09-943-108a-1  
Perfect score: 1368  
Sequence: 1 atggcatttgaggcttacc.....gtatgaattaccggttttaa 1368

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB ID	Description
1	1368	100.0	1368	4	US-09-035-382-1	Sequence 1, Appli
2	790.4	57.8	792	4	US-09-035-382-3	Sequence 3, Appli
3	743.2	54.3	1014	4	US-09-134-001C-1958	Sequence 1958, Ap
4	496	36.3	1569	2	US-08-923-772-1	Sequence 1, Appli
5	496	36.3	1569	4	US-09-385-287-1	Sequence 1, Appli
6	496	36.3	7577	4	US-08-961-527-46	Sequence 46, Appli
7	268	19.6	4804	4	US-09-066-047-6	Sequence 6, Appli
8	252.4	18.5	1098	4	US-09-221-017B-928	Sequence 928, App
9	172.6	12.6	836	4	US-08-858-207A-180	Sequence 180, App
10	145.6	10.6	1251	3	US-09-007-476-1	Sequence 1, Appli
11	144.8	10.6	1248	4	US-09-134-001C-1980	Sequence 1980, Ap
12	127.4	9.3	390	4	US-08-651-155B-4	Sequence 4, Appli
13	102.4	7.5	831	3	US-09-007-484-3	Sequence 3, Appli
14	102.4	7.5	831	4	US-09-309-682-3	Sequence 3, Appli
15	102.4	7.5	1278	2	US-08-986-963-1	Sequence 1, Appli
16	102.4	7.5	1278	2	US-08-986-963-3	Sequence 3, Appli
17	102.4	7.5	2764	2	US-08-986-963-4	Sequence 4, Appli
18	100.8	7.4	1293	3	US-09-007-484-1	Sequence 1, Appli
19	100.8	7.4	1293	4	US-09-309-682-1	Sequence 1, Appli
20	100.8	7.4	13121	4	US-08-961-527-126	Sequence 126, App
21	86.8	6.3	1015	4	US-08-981-527A-7	Sequence 7, Appli
22	86.8	6.3	4370	4	US-08-981-527A-20	Sequence 20, Appli
23	61.6	4.5	7218	1	US-08-232-463-14	Sequence 14, Appli
24	61.2	4.5	2166	2	US-08-317-401E-3	Sequence 3, Appli
25	59.6	4.4	2877	2	US-08-317-401E-1	Sequence 1, Appli
26	58	4.2	29793	4	US-09-302-812-38	Sequence 38, Appli
27	58	4.2	29793	4	US-09-511-477-38	Sequence 38, Appli

C	28	58	4.2	29793	4	US-09-511-507-38	Sequence 38, Appl
	29	52.2	3.8	774	4	US-09-461-697-187	Sequence 187, Appl
	30	52.2	3.8	819	4	US-09-461-697-185	Sequence 185, Appl
	31	52.2	3.8	1669	4	US-09-461-697-184	Sequence 184, Appl
	32	49.4	3.6	1022	4	US-08-960-780-33	Sequence 33, Appl
	33	49.4	3.6	1022	4	US-09-073-898-33	Sequence 33, Appl
	34	48.4	3.5	3279	4	US-08-446-137B-1	Sequence 1, Appl
	35	48.2	3.5	3095	6	5231168-1	Patent No. 5231168
	36	47.8	3.5	4074	1	US-08-471-033-19	Sequence 19, Appl
	37	47.8	3.5	4074	2	US-08-471-044-19	Sequence 19, Appl
	38	47.8	3.5	4074	2	US-08-463-483A-19	Sequence 19, Appl
	39	47.8	3.5	4074	2	US-08-471-046A-19	Sequence 19, Appl
	40	47.8	3.5	4074	2	US-08-470-566B-19	Sequence 19, Appl
	41	47.8	3.5	4074	2	US-08-469-334-19	Sequence 19, Appl
	42	47.8	3.5	4074	3	US-09-300-529-19	Sequence 19, Appl
	43	47.4	3.5	717	4	US-09-461-697-189	Sequence 189, Appl
	44	45.4	3.3	1891	4	US-08-973-462-3	Sequence 3, Appl
	45	45.4	3.3	5361	4	US-08-973-462-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-09-035-382-1  
; Sequence 1, Application US/090355382  
; Patent No. 6284515  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael T.  
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES  
; TITLE OF INVENTION: AND POLYNUCLEOTIDES  
; FILE REFERENCE: GM50035  
; CURRENT APPLICATION NUMBER: US/09/035,382  
; CURRENT FILING DATE: 1998-03-05  
; EARLIER APPLICATION NUMBER: 60/057,890  
; EARLIER FILING DATE: 1997-09-03  
; NUMBER OF SEQ. ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1368  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-035-382-1

Query Match	100.0%;	Score 1368;	DB 4;	Length 1368;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1368;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	ATGGCATTGGAAGCCTTATCAGAACGCTGCAAGCGCATGCAAAAATGCGTGGTAAAG	60	
Db	1	ATGGCATTGGAAGCCTTATCAGAACGCTGCAAGCGCATGCAAAAATGCGTGGTAAAG	60	
QY	61	GGTAAACTTACTGAAGTGATATAAGATAATGCGTGAAGTAAGATTACCGTTATTT	120	
Db	61	GGTAAACTTACTGAAGTGATATAAGATAATGCGTGAAGTAAGATTACCGTTATTT	120	
QY	121	GAGGTGACGTAAACITTAAGTGTGTAAGAAATTTATTAACACATGATCAAGCGGCA	180	
Db	121	GAGGTGACGTAAACITTAAGTGTGTAAGAAATTTATTAACACATGATCAAGCGGCA	180	
QY	181	TTAGGTTCCGATGTAATGCAATCATTACACCGGCAACAAAGTTATTAATAGTCTCAA	240	
Db	181	TTAGGTTCCGATGTAATGCAATCATTACACCGGCAACAAAGTTATTAATAGTCTCAA	240	
QY	241	GATGAATTAACGAAGTTGATGGTGGAGAAATACATCGATTATATGCAATAAACC	300	
Db	241	GATGAATTAACGAAGTTGATGGTGGAGAAATACATCGATTATATGCAATAAACC	300	
QY	301	CCTACTGTTGTATGATGGTGTGTTTACAGTGTGTTAAACACACTCCAGGTAAA	360	
Db	301	CCTACTGTTGTATGATGGTGTGTTTACAGTGTGTTAAACACACTCCAGGTAAA	360	
QY	361	TTAGCATTATTGCGGTAAATAATACAAAAAACCTATGTTAGTTCACGACGATATT	420	

```
|||||
Db 361 TTACGATTTATGTCGTAATAAATACAAACAAACCTGTTAGTTGCAGCAGATATT 420
QY 421 TATCGTCCAGCAGAGTAATCAATTACAAACAGTAGGAAACAAATGTGATTCCTGTA 480
Db 421 TATCGTCCAGCAGAGTAATCAATTACAAACAGTAGGAAACAAATGTGATTCCTGTA 480
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATCAATGATTAACACATGCT 540
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATCAATGATTAACACATGCT 540
QY 541 AAAGAAGAACATTTAGACTTTGTAACTTTGTAATCAATGATTAACACATGCTGATGAA 600
Db 541 AAAGAAGAACATTTAGACTTTGTAACTTTGTAATCAATGATTAACACATGCTGATGAA 600
QY 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAACCAACGAAATTAATGTTA 660
Db 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAACCAACGAAATTAATGTTA 660
QY 661 GTTGTGATTAATCAATGACGGGTCAAGATGCTGTCAATGTTGACAGATCTTTTGACGATCAA 720
Db 661 GTTGTGATTAATCAATGACGGGTCAAGATGCTGTCAATGTTGACAGATCTTTTGACGATCAA 720
QY 721 CTTGATGTCACAGGTGTTACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 780
Db 721 CTTGATGTCACAGGTGTTACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 780
QY 781 TTATCTATTCGTTCCGTCACACAAACCAATTAATTAATTTGTTGTTAGTCAAGTGAAGTTA 840
Db 781 TTATCTATTCGTTCCGTCACACAAACCAATTAATTAATTTGTTGTTAGTCAAGTGAAGTTA 840
QY 841 GATGGTTTACAGCTATTCATCCCTGAACGTTATGCCATCACGTTATTTAGTATGGGTGAT 900
Db 841 GATGGTTTACAGCTATTCATCCCTGAACGTTATGCCATCACGTTATTTAGTATGGGTGAT 900
QY 901 GTGTTAAGTTTAAATGAAAGGCGCAACAGATGTGGATCAAGAAAGAAAGAAAGTTTA 960
Db 901 GTGTTAAGTTTAAATGAAAGGCGCAACAGATGTGGATCAAGAAAGAAAGAAAGTTTA 960
QY 961 GAGAAAGATCGTGAGTCACTGCTGTTTACCTTATGAAATGATTCACAGGTATGAATFAAATG 1020
Db 961 GAGAAAGATCGTGAGTCACTGCTGTTTACCTTATGAAATGATTCACAGGTATGAATFAAATG 1020
QY 1021 GTGAAATCTAGACCACTGATGATTAATGAAATGATTCACAGGTATGAATFAAATG 1080
Db 1021 GTGAAATCTAGACCACTGATGATTAATGAAATGATTCACAGGTATGAATFAAATG 1080
QY 1081 AAAGGCTAGATAAGCTTAATGATGAGTGAAGGAAAGCAAAATTTGATCAATTAAGGATATC 1140
Db 1081 AAAGGCTAGATAAGCTTAATGATGAGTGAAGGAAAGCAAAATTTGATCAATTAAGGATATC 1140
QY 1141 CAGTCAATGACCGGCTGAAGAAACAAATCCAGACACATTAATGATCACTGATCAAGTAAAG 1200
Db 1141 CAGTCAATGACCGGCTGAAGAAACAAATCCAGACACATTAATGATCACTGATCAAGTAAAG 1200
QY 1201 CGTATTCCTAAGGGTCTGGTCTTCATTAAGAAAGTCAATCGTTGATGAACAAATTT 1260
Db 1201 CGTATTCCTAAGGGTCTGGTCTTCATTAAGAAAGTCAATCGTTGATGAACAAATTT 1260
QY 1261 AACGATATGAAGAAATGATGAACAAATTCACCTGGTGGGGTAAAGGTAAGAAAGTTAA 1320
Db 1261 AACGATATGAAGAAATGATGAACAAATTCACCTGGTGGGGTAAAGGTAAGAAAGTTAA 1320
QY 1321 GCATCAATCAATGAATATGTTAAAGGTATGAATTTACCGTTTAA 1368
Db 1321 GCATCAATCAATGAATATGTTAAAGGTATGAATTTACCGTTTAA 1368
```

## RESULT 2

US-09-035-382-3  
; Sequence 3, Application US/09035382  
; Patent No. 6284515  
; GENERAL INFORMATION:

```
; APPLICANT: Black, Michael T.  
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES  
; FILE REFERENCE: GM50035  
; CURRENT APPLICATION NUMBER: US/09/035,382  
; CURRENT FILING DATE: 1998-03-05  
; EARLIER APPLICATION NUMBER: 60/057,890  
; EARLIER FILING DATE: 1997-09-03  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 792  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-035-382-3
```

Query Match 57.88; Score 790.4; DB 4; Length 792;

Best Local Similarity 99.9%; Pred. No. 6.2e-191;

Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGCATTGGAAGGCTTATCAGAACGCCTGCAAGCGACGATGCAAAAAATGCCGTGTAAG 60
Db 1 ATGGCATTGGAAGGCTTATCAGAACGCCTGCAAGCGACGATGCAAAAAATGCCGTGTAAG 60
QY 61 GGTAAACTTACTGAAGCTCATATAAGATAATGATGCGTGAAGTAAGATAGCGTTATTT 120
Db 61 GGTAAACTTACTGAAGCTCATATAAGATAATGATGCGTGAAGTAAGATAGCGTTATTT 120
QY 121 GAGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGGCA 180
Db 121 GAGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGGCA 180
QY 181 TTAGTTCGCGATGTAATCAATCAATTAACACAGGCAACAAAGTTATTAAATAGTTCAA 240
Db 181 TTAGTTCGCGATGTAATCAATCAATTAACACAGGCAACAAAGTTATTAAATAGTTCAA 240
QY 241 GATGAATTAACAGAGTTGATGGTGGAGAAATFACATCGATTAAATGATCAATTAACCA 300
Db 241 GATGAATTAACAGAGTTGATGGTGGAGAAATFACATCGATTAAATGATCAATTAACCA 300
QY 301 CCTACTGTTGTTATGTTGGTTTACAGGTGCTGTTAAACCAACACACGTCAGGTAAA 360
Db 301 CCTACTGTTGTTATGTTGGTTTACAGGTGCTGTTAAACCAACACACGTCAGGTAAA 360
QY 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACCTATGTTAGTTGCAGCAGATTT 420
Db 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACCTATGTTAGTTGCAGCAGATTT 420
QY 421 TATCGTCCAGCAGCGATAAATCAATTAACACAGTAGGGAAACAAATTCATTCCTGTA 480
Db 421 TATCGTCCAGCAGCGATAAATCAATTAACACAGTAGGGAAACAAATTCATTCCTGTA 480
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACCAATTTGTAATGATTAACATGCAATGATGAA 540
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACCAATTTGTAATGATTAACATGCAATGATGAA 540
QY 541 AAAGAAGAACATTTAGACTTTGTAACTTTGTAATCAATGATTAACACATGCTGATGAA 600
Db 541 AAAGAAGAACATTTAGACTTTGTAACTTTGTAATCAATGATTAACACATGCTGATGAA 600
QY 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAACCAACGAAATTAATGTTA 660
Db 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAACCAACGAAATTAATGTTA 660
QY 661 GTTGTGATTAATCAATGACGGGTCAAGATGCTGTCAATGTTGACAGATCTTTTGACGATCAA 720
Db 661 GTTGTGATTAATCAATGACGGGTCAAGATGCTGTCAATGTTGACAGATCTTTTGACGATCAA 720
QY 721 CTTGATGTCACAGGTGTTACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 780
Db 721 CTTGATGTCACAGGTGTTACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 780
QY 781 TTATCTATTCGTTCCGTCACACAAACCAATTAATTAATTTGTTGTTAGTCAAGTGAAGTTA 792
```



```
QY 1 ATGGCAATTGAGGCTTATCAGAACCGCTGCAAGCGAGATGCAAAAATGCGTGTAAAG 60
Db 1 ATGGCAATTGAGGCTTATCAGAACCGCTGCAAGCGAGATGCAAAAATGCGTGTAAAG 60
QY 61 GGTAACTTACTGAGCTGATATAAGATAATGATGCGTGAAGTAAAGTATAGCTTATTT 120
Db 61 GGAATACTTCTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 121 GAGGCTGAGCTAACTTAAAGTGGTAAAGAAATTTATTAACACAGATATCAGAACGCGA 180
Db 121 GAGCGCGAGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
QY 181 TTAGGTTCCGAGTAATGATCAATTAACACAGGCGCAACAGTATTAATAATAGTTCAA 240
Db 181 GTGGGCGATGAGGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 GATGAATTAACGAAGTGTGAGGCTGAGGAGAAATATCATGATTAATATGTCATAAACA 300
Db 241 GAGGAATGACAGCGCTTTAGGTTCTGATACGGCAGAAATATCAAGTCACTAAGATT 300
QY 301 CCTACTGTTGTTATGATGTTGTTTACAGGTGCTGTTAAACCAACCACTGCAGTAA 360
Db 301 CCAACCATCATGATGATGTTGTTTACAGGCTGTTAAACCAACCACTGCTGTA 360
QY 361 TTAGCATTTATGAGCGCTTAAATAACAAACAAACAAACCTGTTAGTTCAGAGATATT 420
Db 361 TTGGCCAAACAACTCAAGAAAGAAATGCTGCTCTTGTGATGATTGCGGCGATATT 420
QY 421 TATCGTCCAGAGCGATTAATCAATTAACAAACAGTAGGAAACAAATGATATTCCTGTA 480
Db 421 TATCGTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 481 TACAGTGAAGAGATCAAGTAAAGCCACAAACAAATGTAATGATGATTAACAAATGCT 540
Db 481 TTGCGACTTGAACAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 AAAGAAAGCAATTTAGACTTTTAACTTAATGATGATGATGATGATGATGATGATGAT 600
Db 541 CAAACTAATCAATCAAGACTATGCTGCTGATGATGATGATGATGATGATGATGATGAT 600
QY 601 GCATTGATGAAGCAATTAAGAAAGTAAAGAAATGCTAAACAAACAAACAAATGATTA 660
Db 601 CTCTCATGAATGAGCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATG 660
QY 661 GTTGTGATTAATGACGGCTCAAGATGCTCAATGTTGAGAACTCTTTGACGATCAA 720
Db 661 GTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
QY 721 CTGATGTCACAGGCTTACCTTAACATAATGATGATGATGATGATGATGATGATGATG 780
Db 721 TTGGAAGTGAAGTGGGCTCATCTTACCAAGATGATGATGATGATGATGATGATGATG 780
QY 781 TTATCTATTCGTCGCTGACACAAACCAATTAATTTGTTGTTGATGATGATGATGAT 840
Db 781 CTGCTGTGTCGATCATCACTGGAACCAACCAATCAAGTTCACGTGACAGTGAAGAA 840
QY 841 GATGCTTTAGAGCTATCCACTGCAAGTATGATGATGATGATGATGATGATGATGATG 900
Db 841 ACAGATATGAACACTTCCACGAGACCTGATGATGATGATGATGATGATGATGATG 900
QY 901 GTGTTAAGTTAATTGAAGAAAGCGCAACAGATGCTGATCAAGAAAGAAAGAAAGTTA 960
Db 901 ATGCTCACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
QY 961 GAGAAAGATGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Db 961 GCTGAGAAGATGCGGGAAGAAACCTTTGATTTAATGATGATGATGATGATGATGATG 1020
QY 1021 GTGAAAATCTAGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Db 1021 GTGAAAATATGGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
QY 1081 AAAGGGCTAGATAAGCTTAATATGATGATGATGATGATGATGATGATGATGATGAT 1140
```

```
Db 1081 CCAGCACTTAAACATGAAGTGGATGAACCCAGATTTGCTGTAACCGTCCATTGG 1140
QY 1141 CAGTCAATGAGCGCGCTGGAAGAAACAATCCAGACACATGAATGATCACTCAAAAAG 1200
Db 1141 TCTTCATGACATCUGAAGACGTGAACCCAGATTTGTTAAATCCAGCCGCTGCCGT 1200
QY 1201 CGTATTGCTAAAGGCTGCTGCTTCACTTACAAAGCAATCTGTTGATGAACAATTT 1260
Db 1201 CGTATTGCTGCTGCTGCTGGAATACATTCGTCGAAGTCAATAAATTCATCAGGACT 1260
QY 1261 AAGATATGACAAATAATGATGAACAATTCATCTGCTGCGGTAAAGTAAAA 1312
Db 1261 AACAGGCTAAACAGCTCATCGAGGCTGTTATGCTCTGGGATATGAATAAAA 1312
```

## RESULT 5

```
US-09-385-287-1
; Sequence 1, Application US/09385287
; Patent No. 6350857
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: NOVEL fih
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,287
; FILING DATE: 30-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/923,772
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd O
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1569 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-385-287-1
```

```
Query Match 36.3%; Score 496; DB 4; Length 1569;
Best Local Similarity 61.1%; Pred. No. 2.3e-116;
Matches 802; Conservative 0; Mismatches 510; Indels 0; Gaps 0;
```

```
QY 1 ATGGCAATTGAGGCTTATCAGAACCGCTGCAAGCGAGATGCAAAAATGCGTGTAAAG 60
Db 1 ATGGCAATTGAGGCTTATCAGAACCGCTGCAAGCGAGATGCAAAAATGCGTGTAAAG 60
QY 61 GGTAACTTACTGAGCTGATATAAGATAATGATGCGTGAAGTAAAGTATAGCTTATTT 120
Db 61 GGAATACTTCTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 121 GAGGCTGAGCTAACTTAAAGTGGTAAAGAAATTTATTAACACAGATATCAGAACGCGCA 180
```



[illegible]

Dd 1201 CQTATGTGCTGGTTCCTGGAAATACATAATTGCTGAGTAACAATAAATTCATCAAGGACTTT 1260

Qy 1261 AACCATATGAAGAAAATGATGAACAATTCACGTGGTGCGGTAAGAAGTAAA 1312  
|||| | ||| | ||| | ||| | ||| |

Dd 1261 ACCAGGCTAACAGCTCATGCAGGCTGTTATGCTCTGGGGATGATGATAAA 1312  
|||| | ||| | ||| | ||| | ||| |

RESULT 6  
US-08-961-527-46/C  
; Sequence 46, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7577 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-961-527-46

Query Match 36.3%; Score 496; DB 4; Length 7577;  
Best Local Similarity 61.1%; Pred. No. 4.2e-116;  
Matches 802; Conservative 0; Mismatches 510; Indels 0; Gaps 0;

QY	1	ATGGCATTTGAGGCTTATATACAGACGCCCTPGONAGCGACGATGCAAAAAAAGCGGTGGTAAAG	60
Db	3033	ATGGCATTTGAAAGTTTAAACAGAACCTTTTCGAGACGCTCTTTTAAAAATCTACGTAAATAAA	2974
QY	61	GGTAAACTTACTGAAGCGTGATATAAGATAATGATGCTGCAAGTAAGATTAGCGTTATTT	120
Db	2973	GGAAAAATCTCGAATCTGATGTCCTCAGAGGCGACCAAGAAATTCGGTTGGCCTTGCTC	2914
QY	121	GAGGCTGAGCTTAACTTTAAAGTGGTAAAGAATTTATTAAACACGATATACAGACGCGCA	180
Db	2913	GAGGCGGACGTTGCGCTGCTGTGTAAAGGACATTATACAGAAAGTTTCGTGACGCTGCA	2854
QY	181	TTAGGTTCCGATGTAAATGCATCATTTAACACACGCGGCAACAGATTATTAAATAATGTTCAA	240
Db	2853	GTGCGGATGAGGCTCATTTGATACATTAACTCTCGGCAACAGATTATTAAATCTTGAT	2794
QY	241	GATGAATTACGAGTTGATGGGTGGAGAAATACATCGCATTAATATGTCAAATAAACC	300
Db	2793	GAGGAATTGACACCGTTTTAGGTTCTGATCGGCAGAAATATCAAGTACACCTTAAGTT	2734



QY 367 TTATTGTCGGTAAATAACACAAACAAACCTATGTAGTTGTCAGCAGATATTATCGT 426  
Db 3855 --ACTGAGTTAAAGAGGATCTTAAACCCGTGTGATGCTGTTAGCATATATCGT 3912  
QY 427 CCAGCAGCATTAATCAATTAACAACAGTACAGGGAACAAATTCATATTCCTGTATACAGT 486  
Db 3913 COTGAGCTCGAGACAGCTGAAGCTTTTGTGCTGATGAGTTGGTATACACAGCTTCCC 3972  
QY 487 GAAGAGATCAAGTAAGCCACAAATTTAATCAATCAATCAATCAATCAATCAATCAAT 546  
Db 3973 ATCCTTGAGGACAAACCACTTGATATTCGAAGCGTCTATGAGGAGCGAGGCTC 4032  
QY 547 GAACATTTAGACTTTTGAATCATTCATACACAGCTGATACACATGATGAGCATTTG 606  
Db 4033 AAAGGACAGATGTTGTTGATGATACAGCGGGCGCTTGATATCAATCAATCAATCAAT 4092  
QY 607 ATGACGAATTAAGAAGTAAGAAATTCCTAAACAAACGAATTTATGTTAGTTGTC 666  
Db 4093 ATAGATGAGCTGAAGTGTGAAGAAGGAGGTATCACACAGCTGAATTTGTTGTTGA 4152  
QY 667 GATTCATGACGGTCAAGTCTGCTCAATGTTGCAGAAATCTTTTGACATCAACTTGAT 726  
Db 4153 GACTCCTTAATGGGCAAGATGCCCTACTATGTCGCAAGTTCAATGAGGAGTTAGGC 4212  
QY 727 GTCAGAGTGTACCTTTAACTAATTAGATGTTGATACAGTGGTGGTGCAGCTTTATCT 786  
Db 4213 ATTACTGGGACGATCTTTACAGGGCGATGTTGATCCTAGGGGTGGTCTATCTGTGCT 4272  
QY 787 AFTCGTTCGGTACACAAACCAATTAATTTGTTGATGAGTGAAGTTAGATGGT 846  
Db 4273 ATGAAGTTGTTGCTGATGCTCTATAAAGTTTCATGCTACGGGAGAGAGCTGAGAT 4332  
QY 847 TTAGAGCTATTCATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 906  
Db 4333 TTGACGATTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4392  
QY 907 AGTTTAAATGAAGGCGCAACAGATCTGATCAAGAAACAAACAAAGATTTAGAGAA 966  
Db 4393 TCTCTTTGAAGAAAGGCGGTAGAGCGGTGGCAAGGATACAAATTAATGAGCTACAGCG 4452  
QY 967 AAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1026  
Db 4453 AAGCCAGAGAGGTAAATTCGATTTGATGATGATGATGATGATGATGATGATGATGAT 4512  
QY 1027 AATCTAGGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1083  
Db 4513 AAAATGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4572  
QY 1084 GGGCTAGATAGCTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143  
Db 4573 CGCAAGTTGGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4632  
QY 1144 TCAATGACGGCGGTGAAGAAACAAACCCAGACATTTGAATGATGATGATGATGATGATGAT 1203  
Db 4633 TCAATGACGAGGAGGAGGCGGATCTCTGATGATGATGATGATGATGATGATGATGAT 4692  
QY 1204 ATTGCTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1263  
Db 4693 ATAGCAAGGCTGCGGAGCTAAAGTTGATGATGATGATGATGATGATGATGATGATGAT 4752  
QY 1264 GATATGAA 1271  
Db 4753 CAGATGAA 4760

RESULT 8  
US-09-221-017B-928/C  
; Sequence 928, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSO for Windows version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P11182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Morroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 928:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1098 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORYPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...1098  
US-09-221-017B-928

Query Match 18.5%; Score 252.4; DB 4; Length 1098;  
Best Local Similarity 53.6%; Pred. No. 8.2e-55;  
Matches 548; Conservative 0; Mismatches 471; Indels 3; Gaps 1;  
QY 55 GGTAAAGGTAACTTACTGAGCTGATATAAAGATAATGATGCTGCTGAGTAAGATTAGCG 114  
Db 1092 GGTGAGGGGAGAAATTACCGAGATCAATGTAGCGAAACTCTCAAGATGTACGTCGTGCC 1033  
QY 115 TTATTGAGGCTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACATATCAGAA 174  
Db 1032 CTTTGGATGCCGAGCTGACTGTCTGTTGCCAAGCAGTTTACGACCTTGGTGAAGGAA 973  
QY 175 CGCGATTAGTTCGGATGTAATGCAATCAATACACAGGCGCAACAGTTATTAATAATA 234  
Db 972 AAAGCCATCGGCCAAACACGTGCTGACCTCCGTCGTCGCGGCGAATCTGAGTGAGATC 913  
QY 235 GTTCAAGATGAATTAACGAGTTGATGGTGGAGAAATACATCGATTATATGTCAAT 294  
Db 912 GTTCATGACGAATTTGGTCTGCTGTGATGGCGACATCTCTGTGGATATTAATCT---CAA 856

QY	295	AAACCACCTACTGTGTGTATGATGAGGTGGTTTACACAGGTGCTGTGTAACAAACACAACTGCA	354
Db	855	GGCAGTCCGGCAGTAATCCCTGATGTCGGGATTCGCAAGGTTCGGGTAAGACCACTTCCTCC	796
QY	355	GGTAATATTAGCACTATTGATGCGCTAAAAAATAACAACAAAAACCTATGTTAGTTGCACGA	414
Db	795	GGCAACTGGCCAAATATGCTCAAGTCCAAAGCAGGCGCAACGACCTCTGCTCGTTCGCTGC	736
QY	415	GATATTATTACGTCACAGCAGGATAAATCAATTAACAACAGTAGGGAACAATAATGATATT	474
Db	735	GACGTGTATCGCCCTCGGGGTATCCACAGCAGCTGATGATCCTCGCGACGACGCTCGGGGTA	676
QY	475	CCTGTATACAGTGAAGGAGATCAAGTAAAGGCCACACAATTTGTAATGATGATTAATAA	534
Db	675	CCCGCTACAGCGACCGGACGACAAAAGCGGTGAGATAGCCCGACACGCAATAGCC	616
QY	535	CATGCTAAAGAAACATTTAGACTTTGTAATCAATTGATACAGCAGGTCCGATTACACATC	594
Db	615	GAAGCCAGACCAAGGGGAAGACGCTTGTCATATCGACACCGCGGTGCTTTGGCCATC	556
QY	595	GATGAACGAPTATGATGAACGAATTAAGAAGTAAAGAAATTTGCTAAACCAACAAAGAATT	654
Db	555	GACGAAGAAATGATCGCGAGATAGAGGCCATCAAGGTCGCATCCAGCGCAACGAAGTC	496
QY	655	ATGTTAGTGTGATCAATCAATCAGCGCTCAAGATGCTGCTCAAGTTGTCAGAACTCTTTTGAC	714
Db	495	CTTTCTGTGTGATCGATGACCGGTACGAGCCGCTGATACAGCCCAAGGAGTTCAAT	436
QY	715	GATCAACTTGATGTCACAGGTGTTAOCTTAACTAAATAGATGCTGATACACGTGCTGTGT	774
Db	435	ACACGCTTGGATTTCSAGGTGTGTCCTTACCAAGCTCGATGAGATACCGGGCGGT	376
QY	775	GCAGCTTTATCTATTCGTTTCGGTGACAAAAAACCAATTAATTTGTTGGTATCACTGAA	834
Db	375	GCOCGCCCTCTGATCCGTCGGTCGTAAACAAGCCCATCAAGTCTGTCGGTACGGGGAG	316
QY	835	AAGTTAGATGGTTAGACTATTCCATCCTGAAGCTATGCGATCACGTTATTTTAGTATG	894
Db	315	AAATGGAAGCAATCATGCTTTACCCCGGAGGTATGGCCGACCGTATCCTTGGCATG	256
QY	895	GGTGATGTTTAAGTTTAATTAAGAAGCGCAACAAGATGGTATCAAGAAAAGCAAAA	954
Db	255	GGGGACATCGATPCGCTGTGGACGTGTCAGGAGCAATACAGGAACGCGGAGATCGC	196
QY	955	GATTTACAGAAAAGHTCCGTGAGTCATCGTTACTTACTAGATGATTTTTAGACAACCTT	1014
Db	195	AACTTGGAGAGAAATAGCCAGAATTCAGTTTCGACCTCAATGACTTCCTTGCOCAGATA	136
QY	1015	GATCAGGTGAAAATCTAGGACCACCTGGATGATATTGAAATATTCATCCAGGTATGAAT	1074
Db	135	CATCAGATCAAAAGATGGSTAACTCAAGAGGTTGCATCCATGATCCCGGAGTGGGA	76
QY	1075	AA 1076	
Db	75	AA 74	

## RESULT 9

US-08-858-207A-180

: sequence 180. Application IIS/08858207A  
03 00 000 207A 100: Sequence 100, Appl. No. 6348328  
: Patent No. 6348328

; CELL NO. 0346320  
: GENERAL INFORMATION:

GENERAL INFORMATION: Black Michael

APPLICANT: BLACK, MICHAEL  
APPLICANT: HODGSON JOHN

APPLICANT: HOAGSON, JOHN  
APPLICANT: KNOWLES, DAVID

APPLICANT: KNOWLES, David

APPLICANT: Nicholas, Richard

APPLICANT: Stodola, Robert

; TITLE OF INVENTION: NO.

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham

STREET: 709 Swedeland

```

/ STATE: PA
/
/ COUNTRY: USA
/ ZIP: 19406-0939
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/858,207A
/ FILING DATE: 09-MAY-1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/017670
/ FILING DATE: 14-MAY-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Gimmi, Edward R
/ REGISTRATION NUMBER: 38,891
/ REFERENCE/DOCKET NUMBER: P50475
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 610-270-4478
/ TELEFAX: 610-270-5090
/ TELEX:
/
/ INFORMATION FOR SEQ ID NO: 180:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 836 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
/ US-08-858-207A-180

```

## RESULT 10

REF ID: A66666  
US-09-007-476-1

US 03 007 470 I  
: sequence 1. Application US/09007476

Sequence 1, Application No. 6759949

Patent No. 6139949  
CENTRAL INFORMATION:

APPLICANT: Black, Michael T.  
TITLE OF INVENTION: NO. 5159949e1 Ptsy  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia

Query Match 10.6%; Score 145.6; DB 3; Length 1251;  
Best Local Similarity 53.6%; Pred. No. 8.9e-28;  
Matches 388; Conservative 0; Mismatches 309; Indels 27; Gaps 3;

QY	276	ATCGATTAAATATGCTCAAAATAAACCCACTTACTGTGTTTATGATGATCGTTTGGTTTACAGGTGC	335
Db	531	AGCTATGAACATAGAGAATGGTCGTTTAAATGTCTATTAAATGTGCGGTGTGAATGGTGT	590
QY	336	TGGTAAACACAACTGCACGGTAAATTAGCATTTATGATGCGCTAAAAAATACAAACAAAA	395
Db	591	TGTTAAACACAACAATTGGAAATTAGCTTACCGATG--TAAATGGAAGGTAAABA	647
QY	396	ACCTATGTTAGTTCACACAGATATTATTCGTCACAGCGGTAATCAATCAATCAAAACAGT	455
Db	648	AGTAATGTTAGTCGGCGGCGATACATTTTAGACGGGTGCTATTGATCAATTTGAAGTGTG	707
QY	456	AGGGAACAAATATGATATTCCTGTACAGTGAAGGAGATCAAGTAAAGGCCAACAAAT	515
Db	708	GGCGCAAGCTGTGGTGTAGATGTAATTAGCCAAAGTAGAGTCTGATCCAGCTGCTGT	767
QY	516	TGTAACATAATGCATTTAAAAACATGCTAAAGAAAGAACATTTTAGACTTTGTATCATTTGATAC	575
Db	768	TATGTATGATGCGATTATGCGCTAAAAACAAAGGTGTGTATATTTTAACTCTGTGATAC	827
QY	576	AGCAGGTGATTAACATTCGATGAAGCATTTGATGAACGGAATTAAGAAAGTAA--	629
Db	828	CGCTGGACGTTTACAAATAAAACAATCTAATGCAGAAATTAGAAAAGTTTAGCGTGT	887
QY	630	-----AGAAATTTGCTAAACCAACCGAAATTTATGTAGTTTTCGATTCAATGAC	677
Db	888	AATTAATCGACAGTGGCCAGATGCGCCTCATGAACGATTAATCATGTTTAGATGCTACAC	947
QY	678	GGGTCAAGATGCTGCTCAATGTTGCAGATCTTTTGGAGATCAACTTGTATGCTACAGGTGT	737
Db	948	TGTTCAAGATGCGTTGTCACAAGCTAGAAACTTTTAAAGAAGTAAACAATTTTACAGTAT	1007

QY	738	TACCTTAACTAAATTAGATGGTGATACACGGTGGTGGTCAGCTTATCTATTCGTCGGT	797
Db	1008	TGTATTAAACCAATTAGATGGTACACCAAAAGGGGTATCGTATTAGCCATTGCGTAATGA	1067
QY	798	GACACAAAACCAATTAATTTGCTGGTATGAGTGAAGCTTAGATGTTTAGAGCTATT	857
Db	1068	ATTGCCATTCCAGTTAAATATGTAGCTTTAGTGAGCCANTTAGATGACCTTACACCAATT	1127
QY	858	CCATCTCAACGCTATGGCATCAGCTATTTTAGTATGGGTGATGGTAAAGTTAAATGA	917
Db	1128	TAAACCTCGAAAGTTATGTCTACGGCTTAT-----TCGCTGATGATGATGAACAAAATGA	1181
QY	918	AAAGCGCACACAGATGTGGATCAGAAAAGACCAAAAGATTTTAGAGAAAGATGCGTGA	977
Db	1182	CGAATTAACACAGTTGAATATGATCAATTTGTACAGAGAAAAGGACGATATCATGG	1241
QY	978	GTCA	981
Db	1242	GSCA	1245

RESULTS. 11

```

US-09-134-001C-1980
; Sequence 1980, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1980
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1980

```

Query Match 10.6%; Score 144.8; DB 4; Length 1248;  
Best Local Similarity 54.6%; Pred. No. 1.4e-27;  
Matches 346; Conservative 0; Mismatches 267; Indels 21; Gaps 2;

QY	276	ATCGATTAAATATCTCAATTAACCACTACTGTTGTATGATGGTTGGTTTACAGGTCG	335
Db	540	AGCAATGAATTTGAAGATGGACGCTTTAAATGCTATACCTGATGTTGGTGTGAATGGTGT	599
QY	336	TGGTAAACAACAACATGCGAGGTAAATTAGCATTTTGGATCGTAAAAAATACAACAATA	395
Db	600	CGGCAAAACACACAAATTTGGTAATTTAGCTTATCGTTATC---ACAAGAAGTAAAAA	656
QY	396	ACCTATGTTAGTTGACGACGATATTTATCGTCCAGCAGCATTAATCAATTACAAACAGT	455
Db	657	AGTAATGTTAGCGCTGGTGATACCTTCAGACCTGGAGCAATTCACAAATTAACGTCG	716
QY	456	AGGGAACAATAATTGATATTCCTGTTATACAGTGAAGGAGATCAAGTAAGGCCAACAAAT	515
Db	717	GGGAAACGTTGGTGGTTGAAGTTGTGAGTCAAAACGAAGTTCTGACCCCTGCAGCAGT	776
QY	516	TGTTACTTAATTCGATTTAAACAATGCTAAAGAACAATTTAGACTTTGTAATCATTTGATC	575
Db	777	AGTATATGATGCGATTAAATGCCGCAAAAATAAAGGCGTAGATATTTAAATTTGTGTATC	836
QY	576	ACGAGGTGCGATTACACATCCGATGACCATGATGAACCATTAAGAAGATTAAGAAGAAAT	635
Db	837	TGCAAGGACGCTTGCAAAATAATCTAATTTAATGCAAGAGTTAGATAAAATGAACCGTGT	896
QY	636	TGCTAA-----ACCAACGAAATTTATGTTAGTTTGCATTCATATGAC	677





```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/986,963
; FILING DATE: December 8, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317/276-3334
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1278 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1278
; US-08-986-963-1

Query Match          7.5%; Score 102.4; DB 2; Length 1278;
Best Local Similarity 51.5%; Pred. No. 7.5e-17;
Matches 300; Conservative 0; Mismatches 261; Indels 21; Gaps 2;

QY 304 ACTGTTGTTATGATGGTTGTTACAGAGTGTGTTAAACAAACAACTGACGAGTAAATTA 363
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 664 ACAGTTATGCTCTTGTGTTGTGAATGTTGTTGGAAACAACTCTCTATCGGAAACTA 723
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 364 GCATTTATTCGCTTAATAAATACACAAACAACTATGTTAGTTCGACGAGATATTTAT 423
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 724 GC---CCACCGGTACAAACAAAGCTGTAAGAAGTCACTGGTGGCAGCAGATACCTTC 780
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 424 CGTCCAGCAGCGATAAATCAATTAACAAACAGTAGGAAACAAATGATATTCTGTATAC 483
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 781 CGTGGGGTGGAGTAGTACAGTACAGTGTGATGGGGCCGACGAGTAGATGTTCCAGTAGTA 840
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 484 AGTGAAGGAGATCAAGTAAAGCCCAACAAATTTGAATGATTAATGATTAACATGCTAAA 543
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 841 ACTGGACCTGAAAGAGCTGATCCAGCCAGCGGTGTTGTTGATGGTATGGAACGTGCCGTG 900
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 544 GAAGACATTTAGACTTTGTAATCATTTGATACAGCAGGTCCGATTACACATCGATGAAGCA 603
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 901 GCTGAAGGTATCGATATTCATGATTGATCTGCTGGTCTGCTGCAAAATAAGGATAAC 960
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 604 TTGATGAACGAATTAAAGAAAGTAAAGAAATTTGCTAA-----ACCA 645
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 961 CTTATGGCTGAGTTGGAAAGATTTGGTTCGTAATTATCAACGTTGTGCCAGAGCAACA 1020
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 646 AAGCAATATATTAGTTGTGATTAATGACGGGTCAAGATGCTGTCATGTTGCAGAA 705
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1021 CATGAACCTTCTTGGCATTGATCATCAACAGTCAAAATGCTCTAGTACAGGCCAAA 1080
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 706 TCTTTTCAGATCAACTTGTATGTCACAGGTGTTACCTTAACATAATAGATGGTGATACA 765
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1081 GAATTTTCGAAATCACACACCTTTACCGGAATTTGTTGACTTAAGATTTGATGGAACTGCT 1140
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 766 CGTGGTGTGACGTTTATCTATTCGTTGCGTGACACAAAAACCAATTAATTTGTTGGT 825
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1141 CGAGAGGTGGTGTCTAGCCATTCGTGAAGAACTCAATATTCCTGTAATAATGATTTGGT 1200
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 826 ATGAGTGAAGATGATGATGGTTAGAGCTATTCATCTCTGAA 867
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1201 TTTGGTGAAGAAATCGATGATATTGGAGAGTTTAACTCAGAA 1242
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: February 25, 2003, 00:20:53  
Job time : 106 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 23:49:06 ; Search time 75 Seconds  
(without alignments)  
10244.303 Million cell updates/sec

Title: US-09-943-108A-1  
Perfect score: 1368  
Sequence: 1 atggcattgagcgttatc.....gtatgaatttaacggttttaa 1368

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications,NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1356.8	99.2	1368	10	US-09-815-242-8968
2	1344	98.2	1368	10	US-09-815-242-8030
3	1341	98.0	1365	10	US-09-815-242-4365
c 4	703.4	51.4	721	10	US-09-815-242-1949
5	568.6	41.6	1416	10	US-09-815-242-3890
6	568.6	41.6	1434	10	US-09-815-242-6764
7	565	41.3	6729	10	US-09-070-927A-37
c 8	516.2	37.7	521	10	US-09-815-242-1985
c 9	516.2	37.7	521	10	US-09-815-242-2013
10	496	36.3	1395	10	US-09-815-242-9427
11	496	36.3	1569	10	US-09-814-041A-1
12	496	36.3	1572	10	US-09-815-242-9156
13	486.4	35.6	1136	10	US-09-974-300-1847
c 14	463.2	33.9	468	10	US-09-815-242-3070
15	443.4	32.4	1389	10	US-09-815-242-6868
16	418.2	30.6	640681	10	US-09-790-988-1
17	370.6	27.1	1362	10	US-09-815-242-6166
18	360	26.3	1306	10	US-09-815-242-9391
c 19	328.4	24.0	330	10	US-09-815-242-2247

c 20	280	20.5	1374	10	US-09-815-242-7850	Sequence 7850, Ap
21	263.6	19.3	292	10	US-09-815-242-3369	Sequence 3369, Ap
22	259	18.9	1641	9	US-09-738-626-2255	Sequence 2255, Ap
23	254	18.6	1347	10	US-09-815-242-7465	Sequence 7465, Ap
24	160.4	11.7	1440	9	US-09-938-842A-877	Sequence 877, App
25	145.6	10.6	1251	10	US-09-815-242-8028	Sequence 8028, Ap
26	143.2	10.5	948	10	US-09-815-242-4244	Sequence 4244, Ap
27	133	9.7	939	10	US-09-815-242-3937	Sequence 3937, Ap
28	133	9.7	1329	10	US-09-815-242-6807	Sequence 6807, Ap
c 29	133	9.7	15363	10	US-09-070-927A-279	Sequence 279, App
30	102.4	7.5	831	10	US-09-827-663-3	Sequence 3, Appli
31	102.4	7.5	1278	10	US-09-815-242-9165	Sequence 9165, Ap
32	100.8	7.4	1293	10	US-09-827-663-1	Sequence 1, Appli
33	97.6	7.1	1290	10	US-09-815-242-9491	Sequence 9491, Ap
c 34	95.6	7.0	643	10	US-09-815-242-2506	Sequence 2506, Ap
35	82.6	6.0	1245	10	US-09-815-242-6997	Sequence 6997, Ap
36	64	4.7	1494	10	US-09-815-242-6281	Sequence 6281, Ap
37	63.8	4.7	405	10	US-09-960-352-13494	Sequence 13494, A
c 38	60.4	4.4	218	10	US-09-815-242-534	Sequence 534, App
c 39	58	4.2	29793	10	US-09-973-451-38	Sequence 38, Appl
40	55.6	4.1	610	9	US-09-736-457-1094	Sequence 1094, Ap
41	55.6	4.1	610	9	US-09-902-941-1094	Sequence 1094, Ap
42	55.6	4.1	610	9	US-09-849-626-1094	Sequence 1094, Ap
43	54.4	4.0	1530	9	US-09-738-626-2259	Sequence 2259, App
44	52.2	3.8	774	10	US-09-922-261-187	Sequence 187, App
45	52.2	3.8	819	10	US-09-922-261-185	Sequence 185, App

## ALIGNMENTS

## RESULT 1

US-09-815-242-8968  
; Sequence 8968, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8968  
; LENGTH: 1368  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1368)  
US-09-815-242-8968

Query Match				99.2%; Score 1356.8; DB 10; Length 1368;			
Best Local Similarity				99.5%; Pred. No. 3.2e-287;			
Matches 1361; Conservative				0; Mismatches 7; Indels 0; Gaps 0;			
Qy	1	ATGCGATTTGAAGGCTTATCAGAACGCTGCGAAGCGAGATGCAAAAATATGCGTGTAAAG	60				
Db	1	ATGCGATTTGAAGGCTTATCAGAACGCTTGCAGCGAGATGCAAAAATATGCGTGTAAAG	60				
Qy	61	GSTAACTTACTGAGCTGATATAGATAATGATGCTGAGTAAGTAAATAGGTTATTT	120				
Db	61	GSTAACTTACTGAGCTGATATAGATAATGATGCTGAGTAAGTAAATAGGTTATTT	120				
Qy	121	GAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGATGATGAGAACGCGCA	180				
Db	121	GAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGATGATGAGAACGCGCA	180				
Qy	181	TTAGGTTCCGATGTAATCAATCAATTAACACAGGCGCAACAGTATTAATAAGTTCAA	240				
Db	181	TTAGGTTCCGATGTAATCAATCAATTAACACAGGCGCAACAGTATTAATAAGTTCAA	240				
Qy	241	GATGAATTAACCAAGCTGATGGGCGGAGAAATACATGATTAATATGTCAAATAAACCA	300				
Db	241	GATGAATTAACCAAGCTGATGGGCGGAGAAATACATGATTAATATGTCAAATAAACCA	300				
Qy	301	CCPACTGTTGTTATGATGTTGTTTACAAAGTGGTGTAAACAAACAACTGAGGTAAA	360				
Db	301	CCPACTGTTGTTATGATGTTGTTTACAAAGTGGTGTAAACAAACAACTGAGGTAAA	360				
Qy	361	TTAGCATTTATGATCGGTAATAATACAAACAAACCTATGTTAGTTGACGAGATAT	420				
Db	361	TTAGCATTTATGATCGGTAATAATACAAACAAACCTATGTTAGTTGACGAGATAT	420				
Qy	421	TATCGTCCAGCGGATTAATCAATTAACAAACAGTAGGGAACAAATGATATTCCTGA	480				
Db	421	TATCGTCCAGCGGATTAATCAATTAACAAACAGTAGGGAACAAATGATATTCCTGA	480				
Qy	481	TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGAATTAATGATTAACAAACAGCT	540				
Db	481	TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGAATTAATGATTAACAAACAGCT	540				
Qy	541	AAAGAGACATTTAGATTTGTAATCATTTGATACAGAGTGCATACACATCGATGAA	600				
Db	541	AAAGAGACATTTAGATTTGTAATCATTTGATACAGAGTGCATACACATCGATGAA	600				
Qy	601	GCATTTGATGAAGCAATTAAGAGAGTAAAGAAATTTGCTAAACCAAGAAATTTATGTTA	660				
Db	601	GCATTTGATGAAGCAATTAAGAGAGTAAAGAAATTTGCTAAACCAAGAAATTTATGTTA	660				
Qy	661	GTTGTCGATTTCAATGACGGTCAAGATGCTGCTCAATGTTGCAAGATCTTTTGACGATCAA	720				
Db	661	GTTGTCGATTTCAATGACGGTCAAGATGCTGCTCAATGTTGCAAGATCTTTTGACGATCAA	720				
Qy	721	CTTGATCTCACAGGTTTACCTTAACAAATTTAGATGTTGATACAGTGGTGGTGCAGCT	780				
Db	721	CTTGATCTCACAGGTTTACCTTAACAAATTTAGATGTTGATACAGTGGTGGTGCAGCT	780				
Qy	781	TTATCTATTCGTTGCTGACACAAAACCAATTAATTTGTTGATGAGTGAAGAAAGTTA	840				
Db	781	TTATCTATTCGTTGCTGACACAAAACCAATTAATTTGTTGATGAGTGAAGAAAGTTA	840				
Qy	841	GATGGTTTAGAGCTATTCCTGCTGAACGATGATGCTATCAGTATTTTAGTATGGGTGAT	900				
Db	841	GATGGTTTAGAGCTATTCCTGCTGAACGATGATGCTATCAGTATTTTAGTATGGGTGAT	900				
Qy	901	GTGTTAGTTTAAATGAAGAGCCGACACAGATGATGATCAACAAAGCAAGAAATTTA	960				
Db	901	GTGTTAGTTTAAATGAAGAGCCGACACAGATGATGATCAACAAAGCAAGAAATTTA	960				
Qy	961	GAGAAAAGATGGTGCAGTCATCGTTTACTTTTATAGATGATTTTATAGAACACTTGATCAG	1020				
Db	961	GAGAAAAGATGGTGCAGTCATCGTTTACTTTTATAGATGATTTTATAGAACACTTGATCAG	1020				

Qy	1021	GTGAAAATCTTAGGACCACTGGATGATATATTATGAAAAATGATTCAGGTATGATAAAATG	1080				
Db	1021	GTGAAAATCTTAGGACCACTGGATGATATATTATGAAAAATGATTCAGGTATGATAAAATG	1080				
Qy	1081	AAAGGCTAGATAAGCTTAAATATGAGTGAAGCAAAATTTGATCATATTAAGCGATATTC	1140				
Db	1081	AAAGGCTAGATAAGCTTAAATATGAGTGAAGCAAAATTTGATCATATTAAGCGATATTC	1140				
Qy	1141	CATCAATCAGCCGGCTGAAAGAAACAATCCAGACACATTTGAATGATCATCACGTAAAAAG	1200				
Db	1141	CATCAATCAGCCGGCTGAAAGAAACAATCCAGACACATTTGAATGATCATCACGTAAAAAG	1200				
Qy	1201	CGTATTCTAAAGGCTCTGTCGTTTCATTTACAGAGCAATTCGTTTGAATCAACAATTT	1260				
Db	1201	CGTATTCTAAAGGCTCTGTCGTTTCATTTACAGAGCAATTCGTTTGAATCAACAATTT	1260				
Qy	1261	AAGCATATGAAGAAATGATGAACAGTTTCACGTGGCGGTAAAGGTAAAAAGGTAAA	1320				
Db	1261	AAGCATATGAAGAAATGATGAACAGTTTCACGTGGCGGTAAAGGTAAAAAGGTAAA	1320				
Qy	1321	CGCAATCAATGCAAAATATGTTAAAAAGGTATGAATTTACCGTTTAA	1368				
Db	1321	CGCAATCAATGCAAAATATGTTAAAAAGGTATGAATTTACCGTTTAA	1368				

RESULT 2

US-09-815-242-8030  
; Sequence 8030, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA. 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8030  
; LENGTH: 1368  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1368)  
US-09-815-242-8030

Query Match 98.2%; Score 1344; DB 10; Length 1368;  
Best Local Similarity 98.9%; Pred. No. 2e-284;  
Matches 1353; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
Qy 1 ATGCAATTTGAAGGCTTATCAGAACGCTGCAAGCGATGCAAAAAAATCGGTGTAAG 60

|||||  
Db 1 ATGGCAATTGAAGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATGCGTGAAG 60  
QY 61 GGTAACTTACTGAAGCTGATTAAGATTAATGATGCGTGAAGTAAAGATTACGTTATTT 120  
Db 61 GGTAACTTACTGAAGCTGATTAAGATTAATGATGCGTGAAGTAAAGATTACGTTATTT 120  
QY 121 GAGGTGACGTAAAGCTTTAAAGTGGTAAAGAAATTTATTAACACATATCAGACGGCA 180  
Db 121 GAGGTGACGTAAAGCTTTAAAGTGGTAAAGAAATTTATTAACACATATCAGACGGCA 180  
QY 181 TTAGTTCGATGTAATCAATCATTAAACAGGCAACAGTTATTAATAATAGTCAAA 240  
Db 181 TTAGTTCGATGTAATCAATCATTAAACAGGCAACAGTTATTAATAATAGTCAAA 240  
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAATATCATCGATTAAATATGTCAAATAACCA 300  
Db 241 GATGAATTAACGAAGTTGATGGTGGAGAAATATCATCGATTAAATATGTCAAATAACCA 300  
QY 301 CCTACTGTTGTTATGATGTTGTTTACAAGGTGCTGTTAAACAACTACAGGTA 360  
Db 301 CCTACTGTTGTTATGATGTTGTTTACAAGGTGCTGTTAAACAACTACAGGTA 360  
QY 361 TTAGCATTTATGATGCTTAAATAATACACAAAAAACCCTATGTTAGTTCGACGATATT 420  
Db 361 TTAGCATTTATGATGCTTAAATAATACACAAAAAACCCTATGTTAGTTCGACGATATT 420  
QY 421 TATCTCCAGCAGGATTAATCAATTAACACAGTAGGAAACAAATTAATGATTCCTGTA 480  
Db 421 TATCTCCAGCAGGATTAATCAATTAACACAGTAGGAAACAAATTAATGATTCCTGTA 480  
QY 481 TACAGTGAAGGATCAAGTAAAGCCACACAAATTTGTAATCAATCAATTAACATGCT 540  
Db 481 TACAGTGAAGGATCAAGTAAAGCCACACAAATTTGTAATCAATCAATTAACATGCT 540  
QY 541 AAAGAAGACATTTAGACTTTTAACTTATGATACAGCAGCTGATACACATCATGATGAA 600  
Db 541 AAAGAAGACATTTAGACTTTTAACTTATGATACAGCAGCTGATACACATCATGATGAA 600  
QY 601 GCATTGATGAATTAAGAAAGTAAAGAAATTTGCTTAACCAACAAATTAATGTTA 660  
Db 601 GCATTGATGAATTAAGAAAGTAAAGAAATTTGCTTAACCAACAAATTAATGTTA 660  
QY 661 GTTGTCGATTCATGAGGCTCAAGATGCTGTCATGTTGCAAGATCTTTTACAGATCAA 720  
Db 661 GTTGTCGATTCATGAGGCTCAAGATGCTGTCATGTTGCAAGATCTTTTACAGATCAA 720  
QY 721 CTGATGTCAGAGTGTACCTTAACTAAATAGATGGTGTATACAGCTGGTGGTCACT 780  
Db 721 CTGATGTCAGAGTGTACCTTAACTAAATAGATGGTGTATACAGCTGGTGGTCACT 780  
QY 781 TTATCTATGCTGCTGAGCAACAAACAAATTAATTTGTTGTAAGTGAAGATTA 840  
Db 781 TTATCTATGCTGCTGAGCAACAAACAAATTAATTTGTTGTAAGTGAAGATTA 840  
QY 841 GATGTTTATGAGCTATTCATCCTGAAGTATGTCATCAAGTATTTAGTATGGTGGAT 900  
Db 841 GATGTTTATGAGCTATTCATCCTGAAGTATGTCATCAAGTATTTAGTATGGTGGAT 900  
QY 901 GTGTTAAGTTTAAATGAAAAAGCGCAACAGATGTTGATCAAGAAAAACCAAGATTTA 960  
Db 901 GTGTTAAGTTTAAATGAAAAAGCGCAACAGATGTTGATCAAGAAAAACCAAGATTTA 960  
QY 961 GAGAAAAAGATGCGTGAAGTATGCTTTTACTTTAGATGATTTTGAACAACTTGATCAG 1020  
Db 961 GAGAAAAAGATGCGTGAAGTATGCTTTTACTTTAGATGATTTTGAACAACTTGATCAG 1020  
QY 1021 GTGAAAAATCTAGGACCACTGATGATTAATTAAGATGTTCCAGGTATGATTAATG 1080  
Db 1021 GTGAAAAATCTAGGACCACTGATGATTAATTAAGATGTTCCAGGTATGATTAATG 1080  
QY 1081 AAAGGCTGATGAAGCTTAATATGATGAAAGCAAAATTCATCATTAATAAGCGATTATC 1140

Db 1081 AAAGTCTAGATGAAGCTTAAATATGATGAAAGCAAAATTCATCATTAATAAGCGATTATC 1140  
QY 1141 CAGTCAATGACCGCGCTGAAGAAACAAATTCACACACATTAATGATATCAGTAAAG 1200  
Db 1141 CAGTCAATGACCGCGCTGAAGAAACAAATTCACACACATTAATGATATCAGTAAAG 1200  
QY 1201 CGTATTGTTAAAGGCTGCTGCTGCTTCAATTAAGAAGTCAATCGTTTGATGAACAATTT 1260  
Db 1201 CGTATTGTTAAAGGCTGCTGCTGCTTCAATTAAGAAGTCAATCGTTTGATGAACAATTT 1260  
QY 1261 AACGATATGAAGAAATGATGAACAAATTCCTGCTGCGGTAAAGSTAAAGAGTAAA 1320  
Db 1261 AACGATATGAAGAAATGATGAACAAATTCCTGCTGCGGTAAAGSTAAAGAGTAAA 1320  
QY 1321 CGCAATCAATGCAAAATATGTTTAAAGGTATGAATTTACCGTTTAA 1368  
Db 1321 CGCAATCAATGCAAAATATGTTTAAAGGTATGAATTTACCGTTTAA 1368

RESULT 3  
US-09-815-242-4365  
; Sequence 4365, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsep, Karl L.  
; APPLICANT: Zysskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011a  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4365  
; LENGTH: 1365  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-4365

Query Match 98.0%; Score 1341; DB 10; Length 1365;  
Best Local Similarity 98.9%; Pred. No. 8.9e-284;  
Matches 1350; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGCGATTGAAGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATGCGTGAAG 60  
Db 1 ATGCGATTGAAGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATGCGTGAAG 60  
QY 61 GGTAACTTACTGAAGCTGATTAAGATTAATGATGCGTGAAGTAAAGATTACGTTATTT 120  
Db 61 GGTAACTTACTGAAGCTGATTAAGATTAATGATGCGTGAAGTAAAGATTACGTTATTT 120  
QY 121 GAGGTGACGTAAAGCTTTAAAGTGGTAAAGAAATTTATTAACACATATCAGACGGCA 180

Db 121 GAGGCTGACGTAACATTTTAAAGTGGTAAAGATTTTATTAACACAGATATCAGAACGCGCA 180  
QY 181 TTAGGTTCCGATGTAATCAATCAATTAACACAGGCGCAACAGTTTATTAATAAGTTCAA 240  
Db 181 TTAGGTTCCGATGTAATCAATCAATTAACACAGGCGCAACAGTTTATTAATAAGTTCAA 240  
QY 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCTCAATAAACCA 300  
Db 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATATATCGATTATATGTCTCAATAAACCA 300  
QY 301 CCTACTGTTGTTATGATGGTGGTTTACAGGTGCTGGTAAACACAACTCGAGGTAAA 360  
Db 301 CCTACTGTTGTTATGATGGTGGTTTACAGGTGCTGGTAAACACAACTCGAGGTAAA 360  
QY 361 TTACATTTATGATCGTAAATAATACACAAAAACCTTATGTTAGTTGGCAGCAGATATT 420  
Db 361 TTACATTTATGATCGTAAATAATACACAAAAACCTTATGTTAGTTGGCAGCAGATATT 420  
QY 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGGAACAAATTTATTCCTGTA 480  
Db 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGGAACAAATTTATTCCTGTA 480  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATGATCAATTAACATGCT 540  
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATGATCAATTAACATGCT 540  
QY 541 AAAGAAACATTTAGCTTTGTAATCATTTGATACAGCAGGTCGATTACATCGATGAA 600  
Db 541 AAAGAAACATTTAGCTTTGTAATCATTTGATACAGCAGGTCGATTACATCGATGAA 600  
QY 601 GCATTGATGAACGAATTAAGAAAGTAAAGAAATTTGCTTAAACCAAGCAATTTATGTTA 660  
Db 601 GCATTGATGAACGAATTAAGAAAGTAAAGAAATTTGCTTAAACCAAGCAATTTATGTTA 660  
QY 661 GTTGTGATTCATCAATGACGGGTCAAGATCGTCAATGTTGCGAAGATCTTTTGACGATCAA 720  
Db 661 GTTGTGATTCATCAATGACGGGTCAAGATCGTCAATGTTGCGAAGATCTTTTGACGATCAA 720  
QY 721 CTGTGATGCACAGGTGTTACCTTAACATAATTAGATGGTGATACACGGTGGTGAGCT 780  
Db 721 CTGTGATGCACAGGTGTTACCTTAACATAATTAGATGGTGATACACGGTGGTGAGCT 780  
QY 781 TTATCTATTGTTGCGGTGACACAAAACCAATTAATTTGTTGTPAGTGAAGTAA 840  
Db 781 TTATCTATTGTTGCGGTGACACAAAACCAATTAATTTGTTGTPAGTGAAGTAA 840  
QY 841 GATGGTTTACAGCTATTCCTGCTGAACGATGCGCATCAGTATTTTGGATGCGGTGAT 900  
Db 841 GATGGTTTACAGCTATTCCTGCTGAACGATGCGCATCAGTATTTTGGATGCGGTGAT 900  
QY 901 GTGTAAAGTTTAAATGAAAAGCGCAACAGATGTGGATCAAGAAAAGCAAAAGATTTA 960  
Db 901 GTGTAAAGTTTAAATGAAAAGCGCAACAGATGTGGATCAAGAAAAGCAAAAGATTTA 960  
QY 961 GAGAAAAGATGCGGTGAGTCATCGTTTACCTTTAGATGATGATTTTGAACAACTTGATCAG 1020  
Db 961 GAGAAAAGATGCGGTGAGTCATCGTTTACCTTTAGATGATGATTTTGAACAACTTGATCAG 1020  
QY 1021 GTGAAAATCTAGGCCACTGGATGATATTTATGAAAATGATTCAGAGTATGAATAAATG 1080  
Db 1021 GTGAAAATCTAGGCCACTGGATGATATTTATGAAAATGATTCAGAGTATGAATAAATG 1080  
QY 1081 AAAGGGCTAGATGAAGCTTAATGAGTGAAAAGCAAAATTCATATTAAGCGGATATC 1140  
Db 1081 AAAGGGCTAGATGAAGCTTAATGAGTGAAAAGCAAAATTCATATTAAGCGGATATC 1140  
QY 1141 CAGTCAATGACCGCGGTGAAGAAAACAAATCCAGACACATTTGAATGATCATCGTAAAAAG 1200  
Db 1141 CAGTCAATGACCGCGGTGAAGAAAACAAATCCAGACACATTTGAATGATCATCGTAAAAAG 1200  
QY 1201 CGTATTGCTAAAGGTCGTGGTTCATTAACAAGAGTCAATCGTTTGATGAACAAATTT 1260  
Db 1201 CGTATTGCTAAAGGTCGTGGTTCATTAACAAGAGTCAATCGTTTGATGAACAAATTT 1260

QY 1261 AACGATATGAAGAAATGATGAACAATTCACGTGGCGGTAAAGGTAAAAAGGTAAA 1320  
Db 1261 AACGATATGAAGAAATGATGAACAATTCACGTGGCGGTAAAGGTAAAAAGGTAAA 1320  
QY 1321 CGCAATCAAAATCAAAATATGTTTAAAGGTATGAATTTACCGTTT 1365  
Db 1321 CGCAATCAAAATCAAAATATGTTTAAAGGTATGAATTTACCGTTT 1365

RESULT 4  
US-09-815-242-1949/c  
; Sequence 1949, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haseibeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1949  
; LENGTH: 721  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-1949

Query Match 51.4%; Score 703.4; DB 10; Length 721;  
Best Local Similarity 99.2%; Pred. No. 1.1e-144;  
Matches 707; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 ATGCATTTGAAGGCTTATCAGAACGCTCGCAACGCGATCAAAAAATCGGTGTAAG 60  
Db 713 ATGCATTTGAAGGCTTATCAGAACGCTTCGCAACGCGATCAAAAAATCGGTGTAAG 654  
QY 61 GGTAACTTACTGAACCTCATATAAGATAATGATCGGTCAAGTAAAGTTAGCGTTATT 120  
Db 653 GGTAACTTACTGAACCTCATATAAGATAATGATCGGTCAAGTAAAGTTAGCGTTATT 594  
QY 121 GAGGCTCAGCTAACTTTTAAAGTGTAAAAAGAAATTTTAAAAACAGTATCAGAACGCGCA 180  
Db 593 GAGGCTCAGCTAACTTTTAAAGTGTAAAAAGAAATTTTAAAAACAGTATCAGAACGCGCA 534  
QY 181 TTAGGTTCCGATGTAATGATGTAATTAACACAGGCGCAACAGTTTATTAATAAGTTCAA 240  
Db 533 TTAGGTTCCGATGTAATGATGTAATTAACACAGGCGCAACAGTTTATTAATAAGTTCAA 474  
QY 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCAATAAACCA 300  
Db 473 GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCAATAAACCA 414

QY 301 CCTACTGTTCTTATGATGGTTGGTTTACAGAGTCTGGTAAACACAACTCCAGGTAAA 360  
DB 413 CCTACTGTTCTTATGATGGTTGGTTTACAGAGTCTGGTAAACACAACTCCAGGTAAA 354  
QY 361 TTAGCATTTATGATGGTAAACACAACTCCAGGTAAACACAACTCCAGGTAAAT 420  
DB 353 TTAGCATTTATGATGGTAAACACAACTCCAGGTAAACACAACTCCAGGTAAAT 294  
QY 421 TATGCTCAGCAGGATTAATCAATTAACAACTAGGGAACAACTTATGATTTCTCTGA 480  
DB 293 TATGCTCAGCAGGATTAATCAATTAACAACTAGGGAACAACTTATGATTTCTCTGA 234  
QY 481 TACAGTGAAGAGATCAAGTAAAGCCACACAACTTGTACTTATCAATTAACAACTGCT 540  
DB 233 TACAGTGAAGAGATCAAGTAAAGCCACACAACTTGTACTTATCAATTAACAACTGCT 174  
QY 541 AAAGAAAGCAATTTAGACTTTTGTATCATTTGATACAGAGTCTGATTAACATCGATGAA 600  
DB 173 AAAGAAAGCAATTTAGACTTTTGTATCATTTGATACAGAGTCTGATTAACATCGATGAA 114  
QY 601 GCATTGATGAAGCAATTTAAAGAGTAAAGAAATGCTTAAACCAACGAAATTTATGTTA 660  
DB 113 GCATTGATGAAGCAATTTAAAGAGTAAAGAAATGCTTAAACCAACGAAATTTATGTTA 54  
QY 661 GTTGTGATTCATGACGGGTCAAGATGCTGTCAATGTTTCAATCTTTTGA 713  
DB 53 GTTGTGATTCATGACGGGTCAAGATGCTGTCAATGTTTCAATCTTTTGA 1

## RESULT 5

US-09-815-242-3890  
; Sequence 3890, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zysskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3890  
; LENGTH: 1416  
; TYPE: DNA  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-3890

Query Match 41.6%; Score 588.6; DB 10; Length 1416;  
Best Local Similarity 64.8%; Pred. No. 3.4e-115;  
Matches 844; Conservative 0; Mismatches 459; Indels 0; Gaps 0;

QY 1 ATGGCATTGTAAGCGTTTATCAGAACCCCTGACAGCAGATGCAAAAAATGCGTGGTAAAG 60  
DB 1 ATGGCTTTTGAAGAGTTTAAACAAACCCCTTACACAGCAGCAATGAGTAAATCCCTCGTAAAG 60  
QY 61 GCTAAACCTTACTGAAGCTGATATAAGAGTAATGATGCGTGAAGTAAAGATTAGATTAGCGTTATTT 120  
DB 61 GGAAGAGTTTCCGAAGCCGACGTTAAAGAAATGATGCGAGAAATCCGTTTGGCTTTATTA 120  
QY 121 GAGGCTGACGTAAATTTAAAGTGTAAAAAGAAATTTATTAACACAGTATCAGAACCGCA 180  
DB 121 GAAGCCGACGTTTAAATTTACAAAGTGTCAAAAGATTTCAAAACCGCTGACAGAACGGCA 180  
QY 181 TTAGTTCCTGATGTAATGCAATCAITTAACACAGGCGCAACAAGTTATTAATAATAGTTCAA 240  
DB 181 GTAGAGTTCGAAGATTTAGAAAGCTTTATCACCAGCCCAACAAATTTGTAATAATTTGTTAT 240  
QY 241 GATGAATTAACGAAGTTGATGGGTGAGAGAAATACATCGATTATATGTCAATTAACCA 300  
DB 241 GAAGAATTAACGAAGCTTTAGGTTTCAGAAACGGTTGAAGTGAATTAATCTCCAAAAATC 300  
QY 301 CCTACTGTTTATGATGTTGGTTTACAAAGTGTGTTAAAAACAACTGCAAGTAAA 360  
DB 301 CCGACAGTGATTTATGATGACAGAGGTTTACAAAGGCGTGTAAACAACTTTACTGTTAA 360  
QY 361 TTAGCATTTATGATGCTTAAAGAAATACAAACAAACCTATGTTAGTTGACAGCATATT 420  
DB 361 TTAGCAAAACACTTTAATGAAAACTGAAACGCTGCTCGCTTTTAAATCGCTGGTGAACCT 420  
QY 421 TATCGTCCAGCAGGATAAATCAATTAACAAACAGTAGGGAACAAATTTGATATTCTCTGA 480  
DB 421 TATCGTCCAGCAGGATTTAGTATGATGTTGAAGGTTTGAAGTCAACAATTAAGATTCCGGT 480  
QY 481 TACAGTGAAGGATCAAGTAAAGCCACAACTTGTAACTTAATGCAATTAACAACTGCT 540  
DB 481 TTTGATATGGACAGATGCTAATCCAGTGGAAATTTGCTCAAGGTTAGCATTAGCA 540  
QY 541 AAAGAAACATTTAGACTTTTGTAACTTATGATACAGCAGTCTGATTAACATCAATGAA 600  
DB 541 AAAGAAACATTTAGACTTTTGTAACTTATGATACAGCAGTCTGATTAACATCAATGAA 600  
QY 601 GCATTGATGAAGAAATTAAGAAAGTAAAGAAATTTGCTAAACCAACAAATTTATGTTA 660  
DB 601 GCTTTAATGGACGAATTAAGAAACAAATTAAGAGTTGGCTAATCCCAATGAATTCGTTA 660  
QY 661 GTTGTGATTCATGACGGGTCAAGATGCTGTCAATGTTGCAAGATCTTTTACAGATCAA 720  
DB 661 GTTGTGATTCATGACGGGTCAAGATGCTGTCAATGTTGCAAGATCTTTTACAGATCAA 720  
QY 721 CTGTGATGTCAGAGTGTACCTTAACATAATTAAGTATGATGATACAGTGTGTCAGCT 780  
DB 721 CTGTGATGTCAGAGTGTACCTTAACATAATTAAGTATGATGATACAGTGTGTCAGCT 780  
QY 781 TTATCTATTTCGTTGACACAAACCAATTAATTTGTTGTTGATGATGATGAAAGTTA 840  
DB 781 CTGTCAATTTCCGGCAGTAAACGGCGCTCCGATTAATTTGTCGGTCTGTTGAAATTA 840  
QY 841 GATGTTTATGAGCTATTCTCACTCTGAACTATGCGATCAAGCTATTTTATGATGATGAT 900  
DB 841 ACCGATTTAGAAATTTTCCATCCGATCGTATGTCGAGTCTGATCTCTAGTATGGGAC 900  
QY 901 GTGTTAAGTTTAATTAAGAAAGCGCAAGATGTTGATCAAGAAACAAACAAATTA 960  
DB 901 ATGTTGACGCTAATTTGAAAGCGCAAGATTTACGATGAGAAACAAACAAATTT 960  
QY 961 GAGAAAGATGCGTGTGATCTATGTTTACTTTTATGATGATTTTATGAAACAACTGATCAG 1020  
DB 961 GCTCAAAATTAAGAAACAGCTTTTGTACCTTAAAGCTTTCAATGAGCAATGATGATCA 1020  
QY 1021 GTCAAAATCTAGGACCACTGGATGATATTATGAAATGATTTCCAGGTATGATGATGAT 1080  
DB 1021 GTTATGGCATGGGACCGGATGAGACTTATTAAAAATGATCCCTGGATGATGATGATG 1080









Db 281 CACCTACTGTGTTATGATGGTTGGTTTACAAGGCTGCTGTAACAAACAACTGAGGTA 222  
QY 359 AATTAGCATTTATCATGCGCTAAATAATACAAACAAACCTATGTTAGTTCAGCAGATA 418  
Db 221 AATTAGCATTTATGATGCGCTAAATAATACAAACAAACCTATGTTAGTTCAGCAGATA 162  
QY 419 TTTATCGTCCAGCAGGATTAATCAATTACAAACAGTAGGGAACAAATTTGATATTCCTG 478  
Db 161 TTTATCGTCCAGCAGGATTAATCAATTACAAACAGTAGGGAACAAATTTGATATTCCTG 102  
QY 479 TATACAGTAGGAGATCAAGTAAGCCACCAACAAATTTGACTAATGATTAACATG 538  
Db 101 TATACAGTAGGAGATCAAGTAAGCCACCAACAAATTTGACTAATGATTAACATG 42  
QY 539 CTAAGAGACATTTAGACTTTGTAATCATGATACAGCA 579  
Db 41 CTAAGAGACATTTAGACTTTGTAATCATGATACAGCA 1

RESULT 10  
US-09-815-242-9427  
; Sequence 9427, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9427  
; LENGTH: 1395  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1395)  
US-09-815-242-9427

Query Match 36.3%; Score 496; DB 10; Length 1395;  
Best Local Similarity 61.1%; Pred. No. 2.4e-99;  
Matches 802; Conservative 0; Mismatches 510; Indels 0; Caps 0;

QY 1 ATGGCATTGAGCTTATCAGACGCTGCAAGCGACGTGCAAAAATGGTGGTAAG 60  
Db 1 ATGGCATTGAGCTTATCAGACGCTTTCGAGAACGCTCTTTAAATAATCTAGCTAAAAA 60  
QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGCGTGAAGTGAAGATGATGCGTTATTT 120  
Db 61 GGTAACTTCTGATCTGATGTCCAAGGCGACCAAGAAATTCGCTTGCCTTGCCTC 120

QY 121 GAGGCTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAAACACATATCAACAGCCCA 180  
Db 121 GAGGCGGACGCTGCTTGGCTTTTAAAGGACTTTATCAAGAAAGTTCGTGAGCGTCA 180  
QY 181 TTAGTCTCCGATGTAATCAATTAACACAGGCAACAAAGTTATTAAATAGTTCAA 240  
Db 181 CTCGGCATGAGGTCATTGATACACTTAATCCTCGGCAACAGATTATTAAATCCTTGAT 240  
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATATACATCGATTAAATATGTCAAAATAACA 300  
Db 241 GAGGAACATGACAGCGCTTTTAGGTTCTGATACGCGAGAAATTTATCAAGTCACCTAAGATT 300  
QY 301 CCTACTGTGTTATGATGTTGTTTACAAGTGTCTGTTAAACAGCACTCCAGGTAAA 360  
Db 301 CCAACCATCATCATGATGTTGTTTACAAGGGCTGGTAAACAACTTTCTGCTGTTAAA 360  
QY 361 TTAGCATTTATTGATGCTGTAATAAATAACAAACAAACCTATGTTAGTTCAGCAGATATT 420  
Db 361 TTGGCCACAACTCAAGAAAGNAGAAATGCTGCTCTTTGATGATTGCGGCGGATATT 420  
QY 421 TATCTCCAGCAGGATAAATCAATTAACAAACAGTAGGGAACAAATTAATGATTCCTCTGA 480  
Db 421 TATCTCCAGCTGCCATTGACCAAGCTTAAGACCTTGGGCAACAGATTGATGCTGCTGC 480  
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCCAACAAATTTGTAACATTAATCCATTAAACATGCT 540  
Db 481 TTTGCACCTTGGAAAGAGATGACCAAGCTTTGAGATTGTAGTCAANGTTGGAGCAAGCC 540  
QY 541 AAAGAAGAACATTTAGACTTTTGAATCATTTAGATACAGCAGTCTGATTACACATCATGATA 600  
Db 541 CAACTAATCAATACAGACTATGCTTTGATGATGATGCTGCTGTTGAGATTGATGAG 600  
QY 601 GCATTGATGAACGAATTAAGAAAGTAAAGAAATGCTTAAACCAACAAATTAATGTTA 660  
Db 601 CTCCTCATGAATGAGCTTCTGATGTGAAAGCAATGCTCAACCAATGAATCTTGCT 660  
QY 661 GTTGTGCAATCAATGACGGGTCAAGATGCTGCTCAATGTTGCAGAAATCTTTTACACATCA 720  
Db 661 GTCTGTGATGCTATGATGTTGTCAGGAAGCAAGCAATGTTGCGGCTGAGTTTAAATCTCAG 720  
QY 721 CTGTGATGTCACAGGTGTTACCTTAATTAATAGATGGTGTATACACGCTGCTGCTGCTGCT 780  
Db 721 TTGGAGTGTACTGGGTCATCTCTACCAGATTGATGCGGATCTGCTGCTGCTGCTGCT 780  
QY 781 TTATCTATTCGTTGCGTGAACAAAACCAATTAATTTGTTGTTGATGATGATAAGTTA 840  
Db 781 CTGCTCTGTTGCTACATTAATGGAACCAATCAAGTTCTACTGCTACAGGTGAAAAGATT 840  
QY 841 GATGTTTATGAGCTATTCATCCTGNAACGATGATGCAATCAGTATTTTAGTATGGGTGAT 900  
Db 841 AGGACATTTGAACCTTCCACCCAGACCGCTGCTAGCCGCTATCCTTGTGTTGGGGAT 900  
QY 901 GTGTTTAACTTTAATTGAAAAGCGCAACAAAGATGTGGATCAAGAAAACCAAGAAATTTA 960  
Db 901 ATGCTCACTTTGATGAGAAAGCTTCTCAGGAATACGATGAACAAAAGCCCTTTGAATG 960  
QY 961 GAGAAAAAGATGCGTGAGTCTGCTTTTACTTTAGATGATTTTGTAGAACAACTTGATCAG 1020  
Db 961 GCTGAGAAAGATGCGGAAAACACCTTTTGATTTTAAATGATTTTCATGATCAATTAAGTACAG 1020  
QY 1021 GTGAAAAATCTAGACCCACTGGATGATATTATGAAAATGATTCAGGTATGATAAATG 1080  
Db 1021 GTGCAAAATATGCGCGCGCTGGAAGACTTGCTCAAGATGATTCAGGTTATGCGGATGCGCAACAAT 1080  
QY 1081 AAAGGCTTAGATGAGCTTTAATATGAGTGAAGCAATTTGATCATATTAAACGGATTATC 1140  
Db 1081 CCAGCAGCTTCAAAACATGAAGGTGGTGAAGCCAGATTGCTCGTAAACGCTGCCAATGTTG 1140  
QY 1141 CAGTCATGACCGCGCTGAAAAGAACCAATCCAGACACATTTGAATGATCATCAATAAAG 1200  
Db 1141 TCTTCGATGACACCTGAGAGACGTGNAACCCAGATTGTTTAATCCAGCGCTGCGCT 1200



; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9156  
; LENGTH: 1572  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1572)  
US-09-815-242-9156

Query Match 36.3%; Score 496; DB 10; Length 1572;

Best Local Similarity 61.1%; Pred. No. 2.5e-99;

Matches 802; Conservative 0; Mismatches 510; Indels 0; Gaps 0;

QY 1 ATGGCATTTGAAGCTTATCAGACGCTGCAAGCGCATGCAAAAATGGTGGTAAG 60  
DB 1 ATGGCATTTGAAGCTTATCAGACGCTGCAAAAATGGTGGTAAG 60  
QY 61 GGTAACTTACTGAAGCTATGAAGATAATGATGCGTGAAGTAAGATTCAGCTTATT 120  
DB 61 GGTAACTTACTGAAGCTATGAAGATAATGATGCGTGAAGTAAGATTCAGCTTATT 120  
QY 121 GAGCTGACGTAACTTTAAAGTGTAAAGAAATTTATTAACACATGACAGCGGCA 180  
DB 121 GAGCTGACGTAACTTTAAAGTGTAAAGAAATTTATTAACACATGACAGCGGCA 180  
QY 181 TTAGCTTCGATGTAATCAATTAACACAGGCGCAACAACTTATTAATAATGTTCAA 240  
DB 181 TTAGCTTCGATGTAATCAATTAACACAGGCGCAACAACTTATTAATAATGTTCAA 240  
QY 241 GATGAATTAAGAGTTGATGGTGGAGAAATATACATCGATTAAATATGCAAAATACCA 300  
DB 241 GATGAATTAAGAGTTGATGGTGGAGAAATATACATCGATTAAATATGCAAAATACCA 300  
QY 301 CTTACTGTTGATGATGTTGTTTACAAAGTGTGTTAAACACAACTGACAGTAA 360  
DB 301 CTTACTGTTGATGATGTTGTTTACAAAGTGTGTTAAACACAACTGACAGTAA 360  
QY 361 TTAGCATTTATGATGCTTAAATAATACAAACAACTATGTTAGTTCACAGCATATT 420  
DB 361 TTAGCATTTATGATGCTTAAATAATACAAACAACTATGTTAGTTCACAGCATATT 420  
QY 421 TATGCTCAGACGATTAATCAATTAACAAAGTAGGAAACAAATTTGATATTCCTGTA 480  
DB 421 TATGCTCAGACGATTAATCAATTAACAAAGTAGGAAACAAATTTGATATTCCTGTA 480  
QY 481 TACAGTGAAGAGATCAAGTAAGCCACAAATTTGTAATTAATGCAATTAACATGCT 540  
DB 481 TACAGTGAAGAGATCAAGTAAGCCACAAATTTGTAATTAATGCAATTAACATGCT 540  
QY 541 AAAGAAGACATTTAGACTTTGATCAATGATACAGAGCTGATTAACATGATGATGAA 600  
DB 541 AAAGAAGACATTTAGACTTTGATCAATGATACAGAGCTGATTAACATGATGATGAA 600  
QY 601 GCATTTGATGACGATTAATTAAGAGTAAGAAATTTGTAATTAATTAATTAATTTA 660  
DB 601 GCATTTGATGACGATTAATTAAGAGTAAGAAATTTGTAATTAATTAATTAATTTA 660

QY 661 GTTGTGATTAATTAAGAGTAAGAAATTTGTAATTAATTAATTTAATGCAATCA 720  
DB 661 GTTGTGATTAATTAAGAGTAAGAAATTTGTAATTAATTAATTTAATGCAATCA 720  
QY 721 CTTGATGTCACAGGTGTACCTTAATTAATTAATTAATTAATTTGTAATGAGTGA 780  
DB 721 CTTGATGTCACAGGTGTACCTTAATTAATTAATTAATTTGTAATGAGTGA 780  
QY 781 TTAATTAATTAATTAATTAATTAATTAATTAATTTGTAATGAGTGAATTAATG 840  
DB 781 TTAATTAATTAATTAATTAATTAATTAATTTGTAATGAGTGAATTAATG 840  
QY 841 GATGTTTATGAGTATTTCCATCTGATGATGATGATGATGATGATGATGATGAT 900  
DB 841 GATGTTTATGAGTATTTCCATCTGATGATGATGATGATGATGATGATGATGAT 900  
QY 901 GTGTTAATTAATTAATTAATTAATTAATTAATTTGTAATGAGTGAATTAATG 960  
DB 901 GTGTTAATTAATTAATTAATTAATTAATTTGTAATGAGTGAATTAATG 960  
QY 961 GAGAAATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
DB 961 GAGAAATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
QY 1021 GTGAAATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
DB 1021 GTGAAATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
QY 1081 AAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
DB 1081 AAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
QY 1141 CAGTCAATGAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
DB 1141 CAGTCAATGAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
QY 1201 CATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
DB 1201 CATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
QY 1261 AACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312  
DB 1261 AACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312

## RESULT 13

US-09-974-300-1847  
; Sequence 1847, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, IB Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE OF INVENTION: Expression  
; FILE REFERENCE: 10085, 500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1847  
; LENGTH: 1136  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-1847

Query Match 35.6%; Score 486.4; DB 10; Length 1136;  
Best Local Similarity 64.3%; Pred. No. 2.8e-97;  
Matches 730; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 1 ATGGCAATTTGAAGGCTTATCAGAACGCTCGAAGCAGCAGATCAAAATATGCGTGTAA 60  
Db 1 ATGGCAATTTGAAGGCTTATCAGAACGCTCGAAGCAGCAGATCAAAATATGCGTGTAA 60  
QY 61 GGTAAATTTACTGAGCTGATATAAGATTAATGTCGCTGAAGTAAGATTAGCGTTATT 120  
Db 61 GGTAAATTTACTGAGCTGATATAAGATTAATGTCGCTGAAGTAAGATTAGCGTTATT 120  
QY 121 GAGCTGACGTAATTTAAAGTGTAAAGATTTATTAACACAGTATCAGAACCGCA 180  
Db 121 GAGCTGACGTAATTTAAAGTGTAAAGATTTATTAACACAGTATCAGAACCGCA 180  
QY 181 TTAGTTCGCGATGTAATCAATCATTAAACACAGGCAACAGTATTAAATAGTTCAA 240  
Db 181 TTAGTTCGCGATGTAATCAATCATTAAACACAGGCAACAGTATTAAATAGTTCAA 240  
QY 241 GATGAATTAACGAATGATGGGTGAGAGAAATACATGATATTAATGTCATTAACCA 300  
Db 241 GATGAATTAACGAATGATGGGTGAGAGAAATACATGATATTAATGTCATTAACCA 300  
QY 301 CCTACTGTTTATGATGTTGTTTACAGGTCGTGTTAAACACACACTGCAGGTAA 360  
Db 301 CCTACTGTTTATGATGTTGTTTACAGGTCGTGTTAAACACACACTGCAGGTAA 360  
QY 361 TTAGCATTATTGATCGTAAAAATACACAAAAAACCCTATGTTAGTTGACAGATATT 420  
Db 361 TTAGCATTATTGATCGTAAAAATACACAAAAAACCCTATGTTAGTTGACAGATATT 420  
QY 421 TATCGTCCAGCAGATTAATCAATTAACACAGTAGGAGAAACAAATGATTCCTGTA 480  
Db 421 TATCGTCCAGCAGATTAATCAATTAACACAGTAGGAGAAACAAATGATTCCTGTA 480  
QY 481 TACAGTGAAGCAGATCAAGTAAACCCACACAAATTTGAATTAATGATTAACACATGCT 540  
Db 481 TACAGTGAAGCAGATCAAGTAAACCCACACAAATTTGAATTAATGATTAACACATGCT 540  
QY 541 AAAGAAACATTTAGCTTTGATCAATCATACACAGCTGATTAACATCATGATGAA 600  
Db 541 AAAGAAACATTTAGCTTTGATCAATCATACACAGCTGATTAACATCATGATGAA 600  
QY 601 GCATGATGAACGAATTAAGAGAGTAAAGAAATTTGCTAAACCAAGAAATTAATGTTA 660  
Db 601 GCATGATGAACGAATTAAGAGAGTAAAGAAATTTGCTAAACCAAGAAATTAATGTTA 660  
QY 661 GTTGTGATCAATGACGGGTCAAGATGCTGTAATGTTGACAGATCTTTGACGATCAA 720  
Db 661 GTTGTGATCAATGACGGGTCAAGATGCTGTAATGTTGACAGATCTTTGACGATCAA 720  
QY 721 CTTGATGTCACAGGTTACCTTAACATAATTAGATGGTGATACACGTTGGTGGCAGCT 780  
Db 721 CTTGATGTCACAGGTTACCTTAACATAATTAGATGGTGATACACGTTGGTGGCAGCT 780  
QY 781 TTAATCATTCGTTGCTGACAAACCAATTAATTTGTTGGTATGAGTGAAGATTA 840  
Db 781 TTAATCATTCGTTGCTGACAAACCAATTAATTTGTTGGTATGAGTGAAGATTA 840  
QY 841 GATGGTTTACAGTATTCATCCTGACAGTATGCAATCAGTATTTTAGTGTGCGTAT 900  
Db 841 GATGGTTTACAGTATTCATCCTGACAGTATGCAATCAGTATTTTAGTGTGCGTAT 900  
QY 901 GTGTTAAGTTTAAATGAAAACGCAACAGATGTGATCAAGAAAACCAAGATTTA 960  
Db 901 GTGTTAAGTTTAAATGAAAACGCAACAGATGTGATCAAGAAAACCAAGATTTA 960  
QY 961 GAGAAAAGATGCTGAGTCAATGCTTTTACTTTAGATGATTTTTTGAACAACTTGATC 1020  
Db 961 GAGAAAAGATGCTGAGTCAATGCTTTTACTTTAGATGATTTTTTGAACAACTTGATC 1020  
QY 1021 GTGAAAATATGACCACTGGATGATTAATGAAATGATTCAGGATGATGAATTAATG 1080  
Db 1021 GTGAAAATATGACCACTGGATGATTAATGAAATGATTCAGGATGATGAATTAATG 1080  
QY 1081 AAGGGCTAGNTAGCTTATATGATGTAAGAGCAAAATTTGATCATTATTAAGCGAT 1136

Db 1081 AAGGGCTTAAAAAAGCTGAAGGTTAGTAAACACAGCTCAGCCATATCGAAGCGAT 1136  
RESULT 14  
US-09-815-242-3070/c  
; Sequence 3070, Application US/09815242  
; Patent No. US20030061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3070  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-3070  
Query Match 33.9%; Score 463.2; DB 10; Length 468;  
Best Local Similarity 99.4%; Pred. No. 2.6e-92;  
Matches 465; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 103 GTAAGATTAGCGTTATTTGAGGCTGACGTAAACTTTTAAAGTGGTAAAAAATTTTATAA 162  
Db 468 GTAAGATTAGCGTTATTTGAGGCTGACGTAAACTTTTAAAGTGGTAAAAAATTTTATAA 409  
QY 163 ACAGTATCAGAACGCCATTAGGTTCCGATGTAATGCAATCAITTAACACCGGACACAA 222  
Db 408 ACAGTATCAGAACGCCATTAGGTTCCGATGTAATGCAATCAITTAACACCGGACACAA 349  
QY 223 GTTATTAATATAGTTCACAGATGAATTAACAGATGATGGTGGAGAAAATACATCGATT 282  
Db 348 GTTATTAATATAGTTCACAGATGAATTAACAGATGATGGTGGAGAAAATACATCGATT 289  
QY 283 AATATGCAAAATAAACCCACTACTGTTTATGATGTTGGTTTACAAAGTCTGGTAAA 342  
Db 288 AATATGCAAAATAAACCCACTACTGTTTATGATGTTGGTTTACAAAGTCTGGTAAA 229  
QY 343 ACAACAACCTCGAGTAAATTAGCATTTAGTATGCGTAAATAATACACAAAAAACCTATG 402  
Db 228 ACAACAACCTCGAGTAAATTAGCATTTAGTATGCGTAAATAATACACAAAAAACCTATG 169  
QY 403 TTAGTTCACACAGATTTATTCGTCACAGCAGGATTAATCAATTAACACAGTAGGAGAA 462  
Db 168 TTAGTTCACACAGATTTATTCGTCACAGCAGGATTAATCAATTAACACAGTAGGAGAA 109  
QY 463 CAAATTGATATTCTCTGATACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAAC 522

```
Db 108 CAAATGATATCTCTGATACAGTGAAGGAGATCAAGTAAAGCCACACAAATGTAAC 49
QY 523 AATGATTAACATGCTAAAGAACATTTAGACTTTGTAATCATT 570
Db 48 AATGATTAACATGCTAAAGAACATTTAGACTTTGTAATCATT 1
RESULT 15
US-09-815-242-6868
; Sequence 6868, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6868
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1389)
US-09-815-242-6868
Query Match 32.4%; Score 443.4; DB 10; Length 1389;
Best Local Similarity 59.2%; Pred. No. 7.2e-88;
Matches 775; Conservative 0; Mismatches 531; Indels 3; Gaps 1;
QY 7 TTTGAGGCTTATCAGACGCGCTGCAGCGACGATGCCAAAATTCGCTGTAAGGTAA 66
Db 4 TTTGAGATTTATCGGATCGCGCTTTCCAAAACCTTACGTAATATCAGAGAAAGCGCT 63
QY 67 CTTACTGAAGCTGATATAAGATAAATGATGCGTGAAGTAAGATTAGCGTTATTGAGGCT 126
Db 64 TTACCGAGAGATAATTAAGAAACCTTACGCGGAGTGCCTATGCGATTAAGGCT 123
QY 127 GAGCTAAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGCGCATTAGT 186
Db 124 GATGTTGCCCTGCGTGGTGAATTTATCGCAAAAAGTAAAGAAAGCGCGCTTGGG 183
QY 187 TCCGATGTAATCATATCAATCAACACAGGCGCAACAGTTATTAATAGTTCAAGTGA 246
Db 184 GAAGAAGTCAATTAAGAAGTTTAACGCCAGGCAAGAATTTTAAAAATCGTTCAAGCGTG 243
QY 247 TTAACGAAGTGTATGGTGGAGAAATATACATCATGATTAATGTAATGTAATTAACCACT 306
```

```
Db 244 CTTGAAAAGCCATGGCGAAGCGAATGAGATTTAAATCTCGCAACCCACACACGCA 303
QY 307 GTTGTATGATGGTGTGTTTACAAAGCTGCTGTAATAACACAACTGCAGGTAAATAGCA 366
Db 304 GTTATCTTAATGCGCGGTTTACAAAGGGCGGTAACACCACAGTGTGGTAAATGGCA 363
QY 367 TTATGTAGTGGTAAATAATACAAACAAAACCTATGTTAGTTGTCAGACAGATATTATCGT 426
Db 364 AAATCTTGGCGTGAACGCCATAAAAGAAAGTGTAGTGGTCTGCTGCTGACGTATATGC 423
QY 427 CCAGCAGCGATAAATCAATTACAAACAGTAGGAAACAAATTTGATTTCTCTATACAGT 486
Db 424 CTTGCTGGGATTAAGCAACTTGAACCTTGGCTCAATCCGTTGGCGTGGATTTTTCACA 483
QY 487 GAAGGAGATCAAGTAAAGCCACACAAATTTGAACATAATCAATTAACAAATGCTAAAGAA 546
Db 484 TCGGATGTTAAACAAACCCCGTTGATTTGTAATCGCGGCTGCTGATGCAAAACTG 543
QY 547 GAACATTTAGACTTTGTAATCATTTAGTACAGAGTGCAGGTGATTAACATTCGATGAAGCAT 606
Db 544 AAATCTACGATGTTGATTTGTTGATAGCGAGTCCGCTACACGTTGATACAGAAATG 603
QY 607 ATGACGGAATTAAGAGAGTAAAGAAATTTGCTAAACCAACAGAAATTAATGTTAGTCTC 666
Db 604 ATGGAGAAATCAAGCAAGTCCATGCTGCATTAATCAATCAATCAATCAATCAATCAAT 663
QY 667 GATTCATAGCGGTCAGAGTGTGTCATTTGTCAGAAATTTTTCGACGATCAACTTGT 726
Db 664 GATGCGATGACTGTCAGAGTGCAGCAATACAGCAAAAGCCCTTAATGAAGCATTCGCT 723
QY 727 GTCACAGGTGTTACCTTAACATAATAGATGATGATACACGTGTCAGTGGTGGTTCATCT 786
Db 724 CTTACAGGGGTATTTTACGAAAAGTGGACGCTGATCGCGCGGTGGTGGCGGTATTCG 783
QY 787 ATTCGCTGGTGCACACAAAACCAATTTAAATTTGTTGATGAGTAAAAAGTTAGATGT 846
Db 784 ATTCTCAATTCACAGGTAAACCAATCAATTTCTTGGGTGGCGGAGAAACAGAGCGG 843
QY 847 TTAGACTATTCATCTCTGAACGATGGCATCACTATTTTATGATGATGGTGAATGTTA 906
Db 844 CTTGAGCCATTCATCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
QY 907 AGTTTAATTAAGAAAGCGCACACAGATGTCGATCAAGAAAGCAAAAGATTTAG--AG 963
Db 904 TCCCTTATCGAAGATCTTGAACGCTTCTGTGATGCTGAAAAGCGGAAAATFGCGCAG 963
QY 964 AAAAGATGCGGTGAGTCACTGTTTACTTTAGATGATTTTATAGAACACTGATCAGGTG 1023
Db 964 AAATTCAGAAAGCGGATGATTTTACTTTAGATGATTTTCCGCGACACAGCTGATCGAA 1023
QY 1024 AAAATCTAGGACCTGATGATATTAAGAAACCAATTCACAGACATTTGAATGATCAACGTA 1083
Db 1024 AAAAATGGCGGATGATGCTATCTGCTTGAAGAAATACCAAGGTGCAAAAATTTGCT 1083
QY 1084 GGCTAGATGAAGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143
Db 1084 GAGCAGTTTAAATAATCAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143
QY 1144 TCAATGACCGCGCTGAAAGAAACCAATTCACAGACATTTGAATGATCAACGTAATAAGCGT 1203
Db 1144 TCCATGACCTTAAAGAACGCTGCGCAACCCAGATATATCAAGAGATCTCCGCTGCTGT 1203
QY 1204 ATTGCTAAAGGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1263
Db 1204 ATTGCTAAAGGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1263
QY 1264 GATATCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312
Db 1264 GAAATGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312
```

Search completed: February 25, 2003, 01:43:52  
Job time : 86 secs

---

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 21:34:56 ; Search time 1861 Seconds  
(without alignments)  
11905.119 Million cell updates/sec

Title: US-09-943-108A-1

Sequence: 1 atggcattgaagcgttacc.....gtatgaattaccgttttaa 1368

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em.estba:\*

2: em.esthum:\*

3: em.estin:\*

4: em.estnu:\*

5: em.estov:\*

6: em.estpl:\*

7: em.estro:\*

8: em.htc:\*

9: gb.estl:\*

10: gb.est2:\*

11: gb.htc:\*

12: gb.est3:\*

13: gb.est4:\*

14: gb.est5:\*

15: em.estfun:\*

16: em.estom:\*

17: gb.gss:\*

18: em.gss.hum:\*

19: em.gss.inv:\*

20: em.gss.pln:\*

21: em.gss.vrt:\*

22: em.gss.fun:\*

23: em.gss.mem:\*

24: em.gss.mus:\*

25: em.gss.other:\*

26: em.gss.pro:\*

27: em.gss.rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	422	30.8	2169	17	BH770912 LLMGtag64
c 2	248.6	18.2	773	13	BF934118 EST554007
c 3	243.8	17.8	711	12	BF940797 7G99A06.x
c 4	232.2	17.0	875	17	AF075878 AF075878
c 5	223	16.3	656	14	B0805434 B0805434
c 6	221.2	16.2	593	10	AW648731 EST327101

7	217.2	15.9	688	17	BH394904
8	210.4	15.4	677	13	BM413418
9	190.6	13.9	605	10	BE432304
10	187.6	13.7	514	13	BM323434
11	186.4	13.6	581	9	AI977909
12	183.4	13.4	814	17	AZ529841
13	179	13.1	1200	11	AY109321
14	178.8	13.1	677	14	BQ96389
15	177.4	13.0	533	10	AW930742
c 16	176.2	12.9	889	17	AZ682189
c 17	163.8	12.0	891	17	BH147501
18	163.2	11.9	2387	11	AK011928
19	158.2	11.6	455	9	AI087737
20	154.2	11.3	554	17	BH369841
21	149.8	11.0	490	13	BM325805
22	148.2	10.8	1032	11	AY106667
23	147	10.7	483	10	BE596266
24	138.6	10.1	476	10	AW033339
25	135.6	9.9	382	10	AW649149
26	133	9.9	1020	17	CNS0780X
27	131.6	9.6	542	13	BF12818
28	127.8	9.3	850	10	AW147928
29	122.6	9.0	899	14	BQ736950
30	120.2	8.8	773	13	BF454259
31	119.6	8.7	703	12	BF936572
32	119	8.7	573	12	EG317280
33	118.6	8.7	655	14	BQ998609
34	118.6	8.7	656	12	BF641333
35	118.6	8.7	1046	13	BF948099
c 36	117.6	8.6	396	9	AJ500175
37	117.6	8.6	791	17	AQ327477
38	117.4	8.6	553	12	BG315860
39	116.6	8.5	307	13	BF307701
c 40	115.8	8.5	481	17	BH376496
41	115	8.4	874	12	BG679799
42	114.2	8.3	610	14	BM944643
43	112.2	8.2	666	9	AI726938
44	112	8.2	547	10	BE461473
45	112	8.2	639	14	BQ608517

#### ALIGNMENTS

RESULT 1  
BH770912/c  
LOCUS  
DEFINITION  
LMMtag644 MG1363 Random Sequence Tag Library Lactococcus lactis  
subsp. cremoris genomic, DNA sequence.  
ACCESSION  
BH770912  
VERSION  
BH770912.1  
KEYWORDS  
GSS.  
SOURCE  
Lactococcus lactis subsp. cremoris  
ORGANISM  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Lactococcus.  
REFERENCE  
1 (bases 1 to 2169)  
Bolotin,A., Ehrlich,S.D. and Sorokin,A.  
TITLE  
Studies of genomes of dairy bacteria Lactococcus lactis  
JOURNAL  
Sci. Aliments, (2002) In press  
COMMENT  
Contact: Sorokin A  
Genetique Microbienne  
INRA  
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France  
Tel: 33 1 34 65 25 16  
Fax: 33 1 34 65 25 21  
Email: sorokine@jouy.inra.fr  
best homologue in strain IL1403 is fffh (91%)  
Class: shotgun  
High quality sequence start: 30  
High quality sequence stop: 2141.  
Location/Qualifiers  
1. .2169

BH770912 2169 bp DNA linear GSS 01-MAY-2002  
LMMtag644 MG1363 Random Sequence Tag Library Lactococcus lactis  
subsp. cremoris genomic, DNA sequence.

ACCESSION  
BH770912  
VERSION  
BH770912.1  
KEYWORDS  
GSS.

SOURCE  
Lactococcus lactis subsp. cremoris

ORGANISM  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Lactococcus.

REFERENCE  
1 (bases 1 to 2169)

Bolotin,A., Ehrlich,S.D. and Sorokin,A.

TITLE  
Studies of genomes of dairy bacteria Lactococcus lactis

JOURNAL  
Sci. Aliments, (2002) In press

COMMENT  
Contact: Sorokin A  
Genetique Microbienne

INRA  
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France

Tel: 33 1 34 65 25 16

Fax: 33 1 34 65 25 21

Email: sorokine@jouy.inra.fr

best homologue in strain IL1403 is fffh (91%)

Class: shotgun

High quality sequence start: 30

High quality sequence stop: 2141.

Location/Qualifiers

1. .2169

```

/organism="Lactococcus lactis subsp. cremoris"
/strain="MGI363"
/db_xref="taxon:1359"
/clone_lib="MGI363 Random Sequence Tag Library"
/notes="Vector: pSGM2; Site_1: SmaI; library of
chromosomal fragments of L.lactis strain MGI363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT      642 a  429 c  380 g   718 t
ORIGIN
Query Match      30.8%; Score 422; DB 17; Length 2169;
Best Local Similarity 63.7%; Pred. No. 1.4e-80;
Matches 689; Conservative 0; Mismatches 390; Indels 3; Gaps 3;
QY  225 TATTAAATAGTTCAGATGAATTAAGCAAGTTGATGGTGGAGAAATACATCGATTA 284
DB  2169 TATCAAAATTTGTATGAAGAATTAACAATATCTTAGCGGTGTGAGCCGGAATTA 2110
QY  285 TATGTCAAATAAACCACTACTGTTGTATGAT-GGTTGGTTTACAGAGTGGTGTAAA 343
DB  2109 CAAATCTCTAAATCCCAACAATATCATGATGGGTTGGTTGCAAGGGCGAGGTAAA 2050
QY  344 CAACAACCTCAGGTAATATAGCATATATGATCGTAAATAATACACAACCAACCTATG- 402
DB  2049 CAACCTTGTCTGTAACTTCACAAAAAAGCTTAAAGAAACAAATGCTCGGCCACTGA 1990
QY  403 TTAGTTGCAGAGATATTTATTCGCCAGCAGCA-TAATCAATTAACAACAGTAGGAA 461
DB  1989 TGATGGCAGCGAGCTTTATCGCCAGCAGCATTTGACCAATTAACCACTGGGAGA 1930
QY  462 ACAATTTGATATCTCTATACAGTGAAGGATCAAGTAAAGCCACACAACAAATGTAAC 521
DB  1929 ACRACTTGACATCTCTGTTTATGATGAAGGAGCTGTGAACCCCGTTAATATTGTCG 1870
QY  522 TAAATGATTAACATGCTTAAAGAAACAATTTAGACTTTGATATCAATTAATGACAGCAG 581
DB  1869 TAATGCTCTTTTAAAGCTCRAAGAAACGTTAAAGATATGTTGTTGATGATGAGCAG 1810
QY  582 TCATTAACATCGATCGATGAAGTATGATGAACGAATTAAGAAAGTAAAGAAATTCCTAA 641
DB  1809 TCGCTTGAATTAATGATACATTTGATGATGATGATGATGATGATGATGATGATGAT 1750
QY  642 ACCAAGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 701
DB  1749 ACCAAGCAATTTTACCTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 1690
QY  702 AGAATCTTTTGACGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 761
DB  1689 TAAATCTTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1630
QY  762 TACACGCGGTGGAGCTTTATCTATCTGCTGGGTGACACAAACCAATTAATTTGT 821
DB  1629 CACACGCGGTGGGTGGGTGATCAATCTGTAATTAACCGGAAATTCAGGAAACCTATTA 1570
QY  822 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 881
DB  1569 AGGACCTGGGATTAATTAACCGACCTGATGATGATGATGATGATGATGATGATGATGAT 1510
QY  882 TATTTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 941
DB  1509 AATCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1450
QY  942 AGAAAAAGCAATTAATTAAGAAAAAGATCGCTGATGATGATGATGATGATGATGATGAT 1001
DB  1449 GGAATCTGCAAACTCGCTGAAAAAATGGCTGATTAACCGTTGATGATGATGATGATGAT 1390
QY  1002 TTTAGAACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1061
DB  1389 GTTGGACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1330
QY  1062 TCCAGGTATGAATTAATGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1121
DB  1329 TCCAGGAATGTCAAAATSCCTGGCTGGATAAAGTTAAAGTTGACCCCTAAAGATGTTGC 1270

```

```

QY  1122 TCATTTAAAGCGATTATCCATCAATGACCGCGCTGAAGAAACAATCCAGACAAATT 1181
DB  1269 TAGAAAACGAGCAATGATTTGTTCAATGACCCCGCTGAACTTTAGAGCAGAACT 1210
QY  1182 GATATTCACGTAAGAAAGCGTATTCCTAAGAGGCTGCTGCTTCATTACAGAAATCAA 1241
DB  1209 ATCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150
QY  1242 TCGTTTGTATGAACAATTTAAGCATATGAGAAATGATGAGAAATTCAGCTGCTGCGG 1301
DB  1149 TAAATTCATTAAGCAATTTAATCAATCCAAAGAAATGATCCAGGAATCAATGAATGTGA 1090
QY  1302 TA 1303
DB  1089 TA 1088

RESULT 2
BI934118 773 bp mRNA linear EST 18-OCT-2001
DEFINITION EST554007 tomato flower, anthesis Lycopersicon esculentum cDNA
LOCUS clone cTOD18J16 5' end, mRNA sequence.
ACCESSION BI934118
VERSION BI934118.1 GI:16248590
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; eusterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 773)
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
Uttterback,I., Van Aken,S., Renning,C.M., Nierman,W., Fraser,C.M.,
Martlb,G.B., Giovannoni,J.J. and Tanksley,S.D.
TITLE Generation of ESTs from tomato flower tissue, anthesis (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
FEATURES
source
1..773
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="tomato flower, anthesis"
/clone_type="flower"
/dev_stage="anthesis"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Cornell University; sequencing: the
Institute for Genomic Research; Flower buds and flowers
were taken from greenhouse plants (4-8 wks old, TA496).
They were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."
BASE COUNT 227 a 129 c 210 g 207 t
ORIGIN
Query Match 18.2%; Score 248.6; DB 13; Length 773;
Best Local Similarity 58.5%; Pred. No. 2.5e-43;
Matches 453; Conservative 0; Mismatches 319; Indels 3; Gaps 1;
QY  116 TATTTAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 175
DB  2 TTTTGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 61
QY  176 GCGATTAGTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 235

```





```

Db 108 AGAAGAGTAGCTAAAGTTCTGGTACACAGTACAGAGNTAA 63

RESULT 4
AF075878/c
LOCUS
DEFINITION
AF075878 Salmonella typhimurium LT2, Lambda DASH II linear GSS 29-AUG-2000
typhimurium genomic clone 245-T3, DNA sequence.
ACCESSION
AF075878
VERSION
AF075878.1
KEYWORDS
GSS.
SOURCE
Salmonella typhimurium.
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1. (bases 1 to 875)
Wong, R.M.-Y., Wong, K.K., Benson, N.R. and McClelland, M.
Sample sequencing of a Salmonella typhimurium LT2 lambda library:
comparison to the Escherichia coli K12 genome
FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
99243757
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifesci.sdsu.edu
Class: shotgun.

FEATURES
source
Location/Qualifiers
1..875
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/clone="245-T3"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
/notice="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 178 a 266 c 217 g 214 t
ORIGIN

Query Match 17.08; Score 232.2; DB 17; Length 875;
Best Local Similarity 56.08; Pred. No. 8.4e-40;
Matches 460; Conservative 0; Mismatches 358; Indels 3; Gaps 1;

QY 296 AACCACTACTGTTTATGATGGTTTACAAAGTCGTGTAACAAACAACTGCAG 355
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 AGCGCCAGCGCTAGTATTGATGCGGTCTGCGAGGGGGTAAACACACAGGTG 772
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 GTAATATGACATATTGTCGTAATAAATACACAAAACCTATGTTAGTTGCAGC 415
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 GTAAGTGGGTAAATCTTGGCGAGAGACACAAAGAAAGTGTGCTGCTCGCG 712
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 416 ATATTTATCGCCAGCAGCAATATCAATTACAAACAGTAGGGAACAAATGATATTC 475
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 AGTCTATCGCCGGCGGATCAACAGCTCGAAGCTGGCTGACAGGTTGGCGTGG 652
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 476 CTGTATACAGTAGAGAGATCAATAAGCCACACAAATTTGTAACATAATGCAATTAAC 535
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 ATTTCTCCCGTCTGATCGGCCAGAAACCGGTTGATGTCACACGCGCTGAAAG 592
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 ATGCTAAGAAGACATTTAGACTTTTATCATTTGATACAGCAGGTCGATTACATCG 595
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 AAGGAAGCTCAAAATCTACAGCTGCTGCTGGTGGATACCGCGGTCTGCTCATGTTG 532
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 596 ATGAAGCATTGTAGCAAGTAATAAAGAGTAAAGAAATTTGTAACCAACCAAGAAATTA 655
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 ACGAAGCGATGATGAGCAAGATCAACACAGTCCACGCTTCATCAACGCCAGTAGAACGC 472
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 556 TGTTAGTTGCAATTCATGAGCGGTCAAGATGCTGTCAATGTTGCGAGATCTTTTGAAG 715
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 TGTTTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 412
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 716 ATCACTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 775
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY	419	TTTATCTGCCAGCAGCAATAATCAATTTCAAAACAGTAGGGAAACAATAATTGATTTCTCTG	478
Db	1	TTTTACAGCCCTGCTGCCATTGATCAACTCACTGACTGCGGTGAACAGTGGGTGTGCCAG	60
QY	479	TATACAGTGAAGGAGATCAAGTAAAGCCACAAACAATTTTAATCAATCAATAAACAATG	538
Db	61	TTTTACTCAGAAGGAACCGCGGCCAAACCTGCAGAAATAACCAAGAATCCGCTGGAAGAG	120
QY	539	CTAAGAGAGACATTTTAGACTTTGTAATCAATGTATACAGCAGCGTCAATACACATCAATG	598
Db	121	CGAAAGAAGATATATTTGATGGCATCTGTAGTGGATCACTGTCGCAGATCTGCAGATTGATA	180
QY	599	AAGCATTTGATGAAGCAATTAAGAAGTAAAGAAGAAATTCGTAAACCAACAAGAAATTAATG	658
Db	181	AAACAATGATGGTTTGAATTGAAGAAGATAAAGAAGGCAGTTAATCTCAGACAGATTGTGC	240
QY	659	TAGTTTGCATTTCAATGACGGGTCAAGATGCTGTCAATGTTCAGAAATCTTTTCAAGCATC	718
Db	241	TTGTTTGTGATGCCATGCTGCCAGGAAGCTGCAGCAATTTGTCACACCTTCAATATTG	300
QY	719	AACTTGATGTCACAGGTGTTACCTTAATCAATTTAGATGGTGATACAGTGCTGGTCCAG	778
Db	301	AAATTTGGTATATCGGTGCTATATTGACTAAATTTGATTTGATTCACAGGGGTGGACGAC	360
QY	779	CTTTATCTATTTCGTTCGGTGCACACAAAACCAATTAATTTGTGTATGATGAGTAAAAAGT	838
Db	361	CACTTAGTGTCAAAAGAGTGTCTGTGAAAAGCCCATCAAGTTTGTGGCGAGGAGAGCGAA	420
QY	839	TAGATGGTTTAGACATATCCATCACTGCACTGATGCGCACTACAGCATTTTAGTATGGGTG	898
Db	421	TGAGGACCTTGACGCTTTCTATCCGACCGCATGGCAACAGCTGTTTGGGAATGGGAG	480
QY	899	ATGTGTTAAGTTTAATTTGAAAAGCGCAACAAGATGTGGATCAAGAAAAGCAAAAGATT	958
Db	481	ATGTCTCTTTCATTTTGTGAAAAGAGCAAGAAGTCGTTCCCAAGAGGATACCATGGAAC	540
QY	959	TAGAGAAGAAAGATCGTGACTCAATCGTTTACTTTTAGATGATTTTTTTAGAACACTTGATC	1018
Db	541	TGCGAGAGAGATCATGAGTCGGAATTTGACTTTCAATGACTTTTTTAAAGCACACACAA	600
QY	1019	AGGTGAAAATCTTAGSACCACCTGGATGATATTTATGAATAATGATTCACAGTATGA	1073
Db	601	ATGTTTGGCAAAATGGGATCCATGAGCCGCTGAATTCGGAAATGATTTCCAGGCATGAA	655



```

/clonelib="cLE663C14"
/clonelib="tomato breaker fruit"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/notes=Vector: pBluescriptSKmCuadapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."

```

```

BASE COUNT      222 a 119 c 165 g 171 t
ORIGIN
Query Match      15.4%; Score 210.4; DB 13; Length 677;
Best Local Similarity 62.1%; Pred. No. 4.1e-35;
Matches 348; Conservative 0; Mismatches 211; Indels 1; Gaps 1;

QY 453 AGTAGGAAACAATTGATTCCTGTATACAGTGAAGGAGATCAAGTAAAGCCACACAA 512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 AGTGGGGAACAGGTTGATGTACCTGTATTCACAGGAGACAGATCAAAACCTGCAGA 150

QY 513 AATTGTACTAATCATTAACATGCTAAAGAGAACATTTAGACTTTGTATTCATTGA 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 AATAGCCGACAGATTACAGAGCCAAAGAGAGATGTAGATGTAGTATATGGA 210

QY 573 TACAGCAGGTGATTACACATCATGATGAAGCATTTGATGAAGCAATTAAGAAAGTAAAGA 632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 TACAGCTGGAAGACTTCAGATAGATAAAACTATGATGATGAATTAAGAGAGTGAACG 270

QY 633 AATTGCTAAACCAACCAATTAATGTAGTTGTGATTCATCAATGACGGGTCAAGATGCTGT 692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 GGTACTGAACCCACAGAGGTTTACTTGTGTGATGCAATGACTGCGCAGAGAGCTGC 330

QY 693 CAATGTTCAGAAATCTTTTACGATCACTTGATGTACAGAGTGTACCTTAACATAAT 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 331 AGCTTTGTGTCACATTCATTCGAATTCGAATTTGATTTGTCCTTCTGACGAGCT 390

QY 753 AGATGGTATACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 AGATGGAGATTTCTAGGGGTGGAGAGCTTTAAGTGTCAAGGAGGTATCAGGAAGCCAA 450

QY 813 TAAATTTGTGTATGATGAAAAGTTAGATGTTTATAGAGTATTCATCTCGAACGAT 872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 CRACTCTAGAGAGGGGTGACGTATGGAGACCTTGAACCTTTCTATCTCGACCGGAT 510

QY 873 GGCATCACATTTTAGG-TATGGGTGATGTTTAAAGTTTAAATGAAAAGCCCAACAG 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 511 GGTGGACGTATTTAGGAATGGGAGATGTTCTATCGTTTGTGAGAAGCCCAAGAA 570

QY 932 ATGTGATCAAGAAAGCAAGATTTAGAGAAAGATCGGTGAGTATCTGTTTACTT 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 571 GTATGGTCAAGAGAGTGTGAAGATTCGAGAGAGATCATGAGTGCAGAAATTTGAT 630

QY 992 TAGATGATTTTATAGAACAA 1011
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 631 TCAATGACTTCTCTGAGCAA 650

```

```

RESULT 9
BE432304
LOCUS      EST398833 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
DEFINITION clone CLE67M2, mRNA sequence.
ACCSSION  BE432304
VERSION   BE432304.1 GI:9430147
KEYWORDS  EST.
SOURCE    tomato.
ORGANISM  Lycopersicon esculentum

```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

```

```

Lycopersicon.
1 (bases 1 to 605)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.

```

```

FEATURES
Location/Qualifiers
1..605
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="CLE67M2"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/notes=Vector: pBluescriptSKmCuadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT      190 a 97 c 159 g 159 t
ORIGIN

```

```

Query Match      13.9%; Score 190.6; DB 10; Length 605;
Best Local Similarity 58.6%; Pred. No. 7.3e-31;
Matches 350; Conservative 0; Mismatches 244; Indels 3; Gaps 1;

QY 192 TCTAATGCAATCATTAACACAGGCAACAGTTATTAAATAGTTCAAGATGAATTAAC 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 TTTGATTCGAGGTGTAACACAGATCAACAACTAGTTAGACTGTCGTCGAGAGCTGT 60

QY 252 GAAGTTGATGGTGGAGAAATATACATCATGATTAATGTCATAAATAAACCCACTGTTGT 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GAACATGATGGTGGAGAGGTTCTGAACTGGTTTGTCTAAATCTAAGCCCCACCGTAAT 120

QY 312 TATGATGGTGGTGGAGAAATATACATCATGATTAATGTCATAAATAAACCCACTGTTGT 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ACTATTCGCGGTCTACAAGGTTTGGGAGACAACTGTTAGTGCAGAGTAGCTTTAT- 179

QY 372 GATGCGTAAATAAATAACAACAAAACCTTATGTTAGTTCGACAGATATTATTCGTCACG 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 --ATCTAAGAGACGCGGTAAAGATTGCTGATGCTGGAGCGGTACAGACCTGC 237

QY 432 AGCGATAATCAATTAACAACAGTAGGGAACAATAATGATATTCCTGTATACAGTAGAGG 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TGCTATTGACCAACTGTTATTATTTGGGTAAACAGGTTGATGATACCTGTTTATGACGAGG 237

QY 492 AGATCAAGTAAAGCCACAAATAATGTTAACTAATGCAATTAATAACATGCTTAAAGAGAAC 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 AACAGATTAACCTTCAGAAATAGCCCGACAGAGATTACAGAGGCCCAAAAGAGAA 357

QY 552 TTTAGACTTTGTAATCATTGATACAGAGGTCGATTACATCATGATGAAGCATTTGATGAA 611
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 TGTAGATTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 417

QY 612 CGAATTAAGAGAGTAAAGAAATTCGTTAAACCAAGAAATATGTTAGTTCGATTC 671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 TGAATTAAGAGAGTAAAGAAATTCGTTAAACCAAGAAATATGTTAGTTCGATTC 477

QY 672 AATGACGGGTCAAGATGCTCAATGTTGCAGATCTTTTGGCGATCAACTGATGTCAC 731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 AATGACGGGTCAAGATGCTCAATGTTGCAGATCTTTTGGCGATCAACTGATGTCAC 537

QY 732 AGGTGTTACCTTAACTAAATTAGATGTTGATACAGTGGTGGTGGTGGTGGTGGTGGTGG 788

```



```
Query Match      13.6%; Score 186.4; DB 9; Length 581;
Best Local Similarity 58.7%; Pred. No. 5.8e-30;
Matches 341; Conservative 0; Mismatches 237; Indels 3; Gaps 1;

QY 111 AGCGTATTATTGAGGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATC 170
Db 3 AGCACTTTTGGAGCCGATAGTATGTTGCCAGTAGTAGAGATTTATTGAGTCTGTAG 62
QY 171 AGAAGCGCATTTAGGTTCCGATTAATGCAATCAATTACACACAGGCGCAACAAGTTATAA 230
Db 63 TGAARAAGGCTGAGCACCGCATGATGATPCCGAGGTCTCGACCTGACGACGATTTGGTAA 122
QY 231 AATAGTTCAGATTAATTAACGAAGTTGATGGGTGGAGAAATACACPCGATTAATATGTC 290
Db 123 GGTGTGTGATGATGAACCTGGTACAACTGATGGCGGGAGGTATPCAGATTTGGTGTTC 182
QY 291 AATATAACACCACTGCTGTGTTATGATGGTTGTTTACAGGTGCTGGTAAACACACAC 350
Db 183 AATATCTGGCCCACTATATATTATTGGCAGGTCTGCAAGGTGTTGGAATAACTACTGT 242
QY 351 TGCAGSTAAATTAGCATTTATGATGCGTAAATAATACACAAAAAACCCTATGTTAGTTGC 410
Db 243 TTCTGCGAAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 299
QY 411 AGCAGATATTATGCTCCAGCAGGATTAATCAATTACAAACAGTAGGGAACAATTTGA 470
Db 300 TGCAGATGTTTACAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 359
QY 471 TATTCCTGTATACAGTGAAGGATCAAGTAAAGCCACAAATTTGTAACATTAATGCAAT 530
Db 360 TGTACCAAGTTACTCAGAGGACTGACGAAACCTTCACAAATAGCCAAACCGTTT 419
QY 531 AAAACATGCTAAGGAAGCAATTTAGACTTTGTAATCATTTGATACAGAGGTGCGATTACA 590
Db 420 GAAGAGGCAAAAGTCAATAGGCTGATGATTAATTTAGTGGACACGCGCTGGAGACTGCA 479
QY 591 CATCGATGAAGCATTTGATGACGAATTAAGAAGTAAAGAAATTCGTAACCAACGA 650
Db 480 GGTAGATTAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
QY 651 AATTATGTTAGTTGCGATTAATGACGGGTCAAGATGCTG 691
Db 540 AGTTCTCTCTGNGGTTGATGCCACTGCTGCGCAAGACG 580

RESULT 12
AZ529841
LOCUS      814 bp      DNA      linear      GSS 03-NOV-2000
DEFINITION Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION  AZ529841
VERSION     AZ529841.1 GI:11083092
KEYWORDS   Entamoeba histolytica.
SOURCE     Entamoeba histolytica
ORGANISM   Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 814)
            Loftus B., Van Aken, S. and Fraser, C.
            Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library
            Unpublished (2000)
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 16
```

```
FEATURES
Source
High quality sequence stop: 811.
Location/Qualifiers
1..814
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
Institute for Genomic Research (TIGR), Rockville, MD
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (1993) Entamoeba histolytica: a
c.g., and Diamond, I.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      342 a      83 c      172 g      217 t
ORIGIN
Query Match      13.4%; Score 183.4; DB 17; Length 814;
Best Local Similarity 54.1%; Pred. No. 2.6e-29;
Matches 373; Conservative 0; Mismatches 316; Indels 0; Gaps 0;
QY 392 AAAAACTTATGTTGTCAGCAGATATTTATCGTCCAGCAGCATATAATCAATTACAAA 451
Db 32 ATAAACCCAGCTGTTGTTGGAGCAGATACATTTAGNCAGCAGCATATGACATTTACAAA 91
QY 452 CAGTAGGGAACAATTTGATTTCCCTGATACAGTGAAGGAGATCAAGTAAAGCCACAC 511
Db 92 TGAATGCTAAACGACGAGGAGTACCATTTTGTGGAATAAAGAAAGAGTATCCAGTAA 151
QY 512 AATTTGTAACATTAATGCAATTAACATGCTTAAAGAAACATTTTAGACTTTGTAATCAT 571
Db 152 AAGTTGCTAGTGAAGAGTTCGTACCTTTAGNAGGAAATAATGATATAATCTTGTGTG 211
QY 572 ATACAGAGGTCGATTTACATCGATGAAGCATTTGATGAACGAATTAAGAAGTAAAG 631
Db 212 ATACATCAGGTAGACATAAACAAGATAAAGAGTTATTTAAAGAAATGCAAGTGTCTGT 271
QY 632 AATTTGCTAAACAAACGAATTTAGTTAGTTCGATTCATTCAGCGGCTCAAGATGCTG 691
Db 272 ATGCAATTAACCCAGATTCATTTATTTTAAATGGATGGAGCTTTTGGACAGCAGCT 331
QY 692 TCAATGTTTGCAGATCTTTTACGATCAACTGATGTCACAGGTGTACCTTTAATAAAT 751
Db 332 TTGGACAAGCAAAAGCATTTTAAAGATGCGATTTGAAGTAGGAAGTGTATTATAACAAA 391
QY 752 TAGATGCTGATACACGTTGGTGGTGGCTTTATCTATTCGTTGGTGCACACAAAACCAA 811
Db 392 TAGATGGACATCAAAATGGAGGAGGAGCATTAAGTGTCTGTAGCAGCAACAAAAGTCCA 451
QY 812 TTAATAATTTGTTGATGAGTGAAGATTTAGATGTTTGAAGACTATTCCTCTGAACTGA 871
Db 452 TTATATTATTGGACAGGAGAAAGTCAATGAATAGAGATTTGTGACAGAAAGTT 511
QY 872 TGCAATCAGTATTTTAGTGTATGGGTGATGTTTAAAGTTTAAATGAAAAGCGCAACAG 931
Db 512 TTGTAAGAAAAATTTACTTTGGAAATGGGAGATTTAAAGGAATAGCAAAATAGCTAAAG 571
QY 932 ATGTCGATCAAGAAAAGCAAAAGATTTAGACAAAAGATCGCTGAGTTCATGTTTACT 991
Db 572 TTGAGAGAAATGCTGAATATATAAACTTTGGTAAACATCTCTCAAGAGAGTCAATTAC 631
QY 992 TAGATGATTTTTTGAACAACATTTGATCAGGTGAAAAATCTAGGACCATGGATGATTA 1051
Db 632 TAGAGATTTGGAAGACAAATTTCAATTTACAAAAAATGGGACATTTAGGAATATAA 691
QY 1052 TGAATAATGATTCAGGTATGAATATAATG 1080
Db 1052 TGAATAATGATTCAGGTATGAATATAATG 1080
```







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 00:20:56 ; Search time 3139 Seconds  
(without alignments)  
4218.470 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274

Sequence: 1 MAFEGSELRQATQMKMRGK.....GKKGRNQMNLMKGMNLPF 455

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV-xlp  
-Q/cgn2\_1/USPTO.spool/US09943108/runat\_21022003\_151630\_1801/app.query.fasta\_1.647  
-DB=GenEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LISN=45  
-DOCALIGN=200 -THR SCORE=oct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09943108 -CGN\_1\_1\_3745 -runat\_21022003\_151630\_1801 -NCP0=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sv.\*  
13: gb\_un.\*  
14: gb\_vt.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rtd.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgt\_hum.\*  
40: em\_hgt\_mus.\*  
41: em\_hgt\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2274	100.0	1368	6	E36050 Signal reco
2	2274	100.0	303750	1	AP003133 Staphyloc
3	2274	100.0	347235	1	AP003361 Staphyloc
4	2261	99.4	295350	1	AP004826 Staphyloc
5	1719.5	75.6	2256	6	AX416333 Sequence
6	1717.5	75.5	347050	1	AL591981 Listeria
7	1708.5	75.1	313450	1	AL596170 Listeria
8	1708.5	75.1	319630	6	AX413016 Sequence
9	1708.5	75.1	349980	6	AX417046 Sequence
10	1647	72.4	2739	1	DI4356
11	1647	72.4	208780	1	BSUB00009
12	1584	69.7	303249	1	AP001515 Bacillus
13	1402	61.7	1136	6	AX433432 Sequence
14	1393	61.3	11024	1	AE007684 Clostridi
15	1382	60.8	296750	1	AP003191 Clostridi
16	1352.5	59.5	14571	1	AE013104 Thermococ
17	1315	57.8	11023	1	AE006391 Lactococ
18	1312	57.7	1605	6	AX194269 Sequence
19	1307	57.5	10462	1	AE008489 Streptoc
20	1306	57.4	1572	6	AX194043 Sequence
21	1306	57.4	7577	6	BD003726 Polynucle
22	1306	57.4	9985	1	AE007428 Streptoc
23	1305	57.4	792	6	E36051 Signal reco
24	1299	57.1	77743	2	SPNEU1910
25	1297	57.0	1569	6	AR081886 Sequence
26	1297	57.0	1569	6	AR195303 Sequence
27	1297	57.0	1569	6	E35383 Novel fph
28	1286	56.6	4152	1	U88582 Streptococ
29	1285	56.5	50946	1	AE014153 Streptococ
30	1283	56.4	10173	1	AE010040 Streptoc
31	1279.5	56.3	12187	1	AE006560 Streptoc
32	1238	54.9	11825	1	AE010644 Fusobacte
33	1180.5	51.9	15721	1	AE001802 Thermotog
34	1173.5	51.6	21387	1	AE008821 Salmonell
35	1173.5	51.6	274050	1	AL627276 Salmonell
36	1168	51.4	1641	6	AX122339 Sequence
37	1168	51.4	337200	1	AP005280 Corynebac
38	1168	51.4	349980	6	AX127150 Sequence
39	1164.5	51.2	4586	1	ECTRMD X01818 E. coli trm
40	1164.5	51.2	10267	1	AE005491 Escherich
41	1164.5	51.2	11272	1	AE000347 Escherich
42	1164.5	51.2	270365	1	AP002562 Escherich
43	1150.5	50.6	10566	1	AE004142 Vibrio ch
44	1150	50.6	11506	1	AE004793 Pseudomon
45	1139.5	50.1	11004	1	AE013692 Yersinia

# ALIGNMENTS

RESULT 1

E36050  
 LOCUS 1368 bp DNA linear PAT 18-JUN-2001  
 DEFINITION Signal recognition particle polypeptide and polynucleotide.  
 ACCESSION E36050  
 VERSION E36050.1 GI:13022452  
 KEYWORDS JP 1999235183-A/1.  
 SOURCE unidentified.  
 ORGANISM unclassified.

## REFERENCE

1. (bases 1 to 1368)  
 Michael, T.B.  
 Signal recognition particle polypeptide and polynucleotide  
 TITLE Patent: JP 1999235183-A 1 31-AUG-1999;  
 JOURNAL SMITHKLINE BEECHAM CORP

## COMMENT

OS Unidentified  
 PN JP 1999235183-A/1  
 PD 31-AUG-1999  
 PF 03-SEP-1998 JP 1998289963  
 PR 03-SEP-1997 US 60/057890, 05-MAR-1998 US 09/035382 PI  
 MICHAEL TERENCE BLACK  
 PC C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00,  
 A61K31/00,  
 PC A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/70, A61K35/76,  
 A61K39/085,  
 PC A61K39/395, A61K39/395, A61K48/00, C07K14/31, C07K16/12, C12N1/15,  
 C12N1/19,  
 PC C12N1/21, C12N5/10, C12N5/10, C12P21/02, C12Q1/68, G01N33/15, PC  
 G01N33/566//  
 PC C12P21/08, G01N33/577, (C12N15/09, C12R1.445), C12N15/00, C12N5/00,  
 C12N5/00,  
 PC (C12N15/00, C12R1.445)  
 CC Strandedness: Single;  
 CC Topology: Linear;  
 PH Key Location/Qualifiers  
 FT source 1..1368  
 FT Location/Qualifiers  
 source 1..1368  
 /organism="Unidentified".  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

## FEATURES

BASE COUNT 521 a 186 c 288 g 373 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 6,44e-130 Length: 1368  
 Score: 2274.00 Matches: 455  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-943-108A-2 (1-455) x E36050 (1-1368)

QY 1 MetAlaPheGluClyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 DB 1 ATGGCATTTGAAGGCTTATCAGACGGCTGCAAGCGACGATGCAAAAAATCGTGTGAAG 60  
 QY 21 GlyLysLeuThrGluAlaAspPheLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 DB 61 GGTAAACTTACTGAAGCTGATATAAGATATGATGCGTGAAGTAAAGATTAGCGCTTATT 120  
 QY 41 GluAlaAspValAsnPhelysValValLysGluPheIleLysThrValSerGluArgAla 60  
 DB 121 GAGGCTGACGTAAACTTAAAGTGTAAAGATTTATTAAACACAGTATCAGACGCGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 DB 181 TTAGGTTCCGATGTAATGCAATCATTAACACCGGGCAACAGTATTAAAAATAGTCTCA 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 DB 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTAAATATGCAATAAACCA 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 DB 301 CCTACTGTTGTTATGATGTTGGTTTACAAGGTCGCGTAAACAAACAACACTGCAGGTAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTytrAsnLysLysProMetLeuValAlaAlaAspIle 140  
 DB 361 TTAGCATTTATGATCGGTAAAAAATACACAAAAAACCTATGTTAGTTGCGAGCATATT 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 DB 421 TATGCTCCAGCAGCATTAATCAATTACAACAGTAGGAGAACAAATGTATATCTCTGTA 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180  
 DB 481 TACAGTGAAGGAGATCAAGTAAACCCACAAATGTGAATGATGCAATTAACACATGCT 540  
 QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
 DB 541 AAAAGAAACATTTAGACTTTGTAACTCATTGATACACAGGTCGATTACACATCGATGAA 600  
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 DB 601 GCATTGATGACGAATTAAGAAGAGTAAAGAAATGCTTAACCAACAGCAATATATGTTA 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
 DB 661 GTTCTCGATTCAATGACGGGTCAAGATGCTGTCAATGTTGCGAATCTTTTGACGATCAA 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 DB 721 CTTGATGTCACAGGTGTTACCTTAACATAATAGATGCTGATACACGTTGGTGGTCAGCT 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 DB 781 TTAATCTATTCGTCGGTGACACAAAAACCAATTAATTTGTTGTTAGCGAGAAAAGTTA 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 DB 841 GATGGTTTAGAGCTATTCCATCCTGAACTGATGCGATCAGCTATTTTAGTATGGGTGAT 900  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 DB 901 GTGTTAAGTTTAATTGAATAAGCGCAACAGATGTGGATCAGAAAAAGCAAGATTTA 960  
 QY 321 GlnLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340  
 DB 961 GAGAAAAAGATCGGTGAGTCATCGTTTACTTTTAGATGATTTTGTAGAACAACTTGATCAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 DB 1021 GTGAAAAATCTAGGACCACCTGGATGATATTATGAAATGATTCCAGGTATGAATAAATG 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspPheIleLysAlaIleIle 380  
 DB 1081 AAAGGCTAGATAAGCTTAATATGAGTGAAGCAAAATGATCATATTAAAGCGCATATC 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400  
 DB 1141 CAGTCAATGACCGCGCTGAAAGAAACAAATCCAGACACATGGAATGATATCAGCTAAAAAG 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
 DB 1201 CGTATTGCTAAAAAGGTCGTGCTGTTTCAATTAACAAGAGTCAATCGCTTTGATGAAACAATTT 1260  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440  
 DB 1261 AACCATATGAAGAAATGATGAAACAAATTCACGTGTCGGGTAAAGGTAAAAAGGTAAA 1320  
 QY 441 ArgAsnGlnMetGlnAsnMetLysLysGlyMetAsnLeuProPhe 455  
 DB 1321 CGCAATCAAAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365

RESULT 2

```

AP003133      303750 bp   DNA   linear   BCT 02-JUL-2002
LOCUS
DEFINITION    Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
genome, section 5/10.
ACCESSION    AP003133 BA000018
VERSION      AP003133.2 GI:14349175
KEYWORDS
SOURCE       Staphylococcus aureus subsp. aureus N315 (sub-species:aureus N315,
strain:N315) DNA.
ORGANISM     Staphylococcus aureus subsp. aureus N315
REFERENCE    1 Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS      Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Iian,J., Ito,T., Kanamori,M.,
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,
Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,
Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
TITLE        Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL      Lancet 357 (9264), 1225-1240 (2001)
MEDLINE      21311952
PUBMED       11418146
REFERENCE    2 (bases 1 to 303750)
AUTHORS      Aoki,K., Oguchi,A., Hosoyama,A., Nagai,Y., Kuroda,M., Hiramatsu,K.
and Kikuchi,H.
TITLE        Direct Submission
JOURNAL      Submitted (30-JAN-2001) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchinite.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
COMMENT      On Jun 12, 2001 this sequence version replaced gi:13701012.
FEATURES     Location/Qualifiers
             1. 303750
             /organism="Staphylococcus aureus subsp. aureus N315"
             /strain="N315"
             /sub_species="aureus N315"
             /db_xref="taxon:158879"
             202..1155
             /gene="SA1056"
             202..1155
             /gene="SA1056"
             /note="ORFID:SA1056"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB42308.1"
             /db_xref="GI:13701013"
             /translation="MKTIGCLLMLLVAGSGGNHKLSSKKSSEKQFTVKES
             EEKDPDLKYEIEKKMKGIKADSLDKLDPMEKSFTHSKGICQWKYKELGKY
             ELADYFTKDSGSSIKDYDAFFPKGKGIKRVYIEHDDVQVDYIWVDPGKKIGNS
             NTPVFAEINTKIDKGLVYASVEPGSIVYHKDDAKYDDSKLKLSQLTKLDHFKP
             VPVSAQIKSFQVPLTSVTFMTGSKDTKDEVLPAAYTFSPKNYEDKSNFDPKVLN
             LVAMDINASSDFGNAHFVLSKYIKESYNSVETASDSSLK"
             /complement(1318..1521)
             /gene="SA1057"
             complement(1318..1521)
             /note="ORFID:SA1057"
             /codon_start=1
             /transl_table=11
             /product="conserved hypothetical protein"
             /protein_id="BAB42309.1"
             /db_xref="GI:13701014"
             /translation="MQVNVKIYIILLFLGSGFIHKFYAGKMQGILHIFCWTGIPH
             ILAIIASAVITVFKPADEQGNVTL"
             /complement(1820..2308)
             /gene="SA1058"
             1820..2308
             /gene="SA1058"
             /note="ORFID:SA1058"

```

```

hypothetical protein, similar to polypeptide deformylase"
/codon_start=1
/transl_table=11
/protein_id="BAB42310.1"
/db_xref="GI:13701015"
/translation="MAIKLVPASHPIILTKAAQAVKTFDDSLKRLLODLEDYMAQEA
AGLCAPQINQISQVAIIDMEGLQJLNPKIISQSNETITDLEGSITLDPYVIGVTR
SKMIVSYDNGKVELTAHDVARMILHIDQMGIPPTERADRLITDKVEAYFI
NP"
2301..3236
/gene="SA1059"
2301..3236
/gene="SA1059"
/note="ORFID:SA1059"
/codon_start=1
/transl_table=11
/product="methionyl-tRNA formyltransferase"
/protein_id="BAB42311.1"
/db_xref="GI:13701016"
/translation="MTKILIFMGTPDESTTVLEMLIAEDVIAVVTQDPVPGRKRVMT
PPVKKVAMKYDLPVYQPEKLSGESELEQLQLDVLIVTAFGQLLPESLIALPKLG
AINVHASLLPKYRGAPIHQAIIDGQETGITIMYVKILDAGNIISQAKIENBDN
VGTMDKLSVLGADLLKETPLPSIIEGTNESVPQDQTQATFASNIREDERISWKEPGR
OVFNQIRGLSPWPVAVTMDDTNLYKIYDAELVETKINERPTIETTKAIIIVATNDN
EAVAKDMLQLAGKKRMLAANYLSGAQNTLVGKKLI"
3155..4540
/gene="SA1060"
3155..4540
/gene="SA1060"
/note="ORFID:SA1060"
hypothetical protein, similar to RNA-binding Sun protein"
/codon_start=1
/transl_table=11
/protein_id="BAB42312.1"
/db_xref="GI:13701017"
/translation="MSWEKENVSCQLFKWCAKHTSRBETMIENVRSLAFDTIQILN
EGAYSNLRINEVSENLNAMDALFTIVYGVTVKRYTLDVLKPVVTKIKAWVRQ
LLWMSIYOYVLDKVPNHAIINEAVEIAKERGGYHNGVNGTILRTMWSDLDPFNEI
ADPKRMAIEYSNEKWIIDWHWHYGLETEITLQSFLETSTTVRANITRASLDRII
EKLQDEIDVEKDHDPYCUHIGGQPIIHSRSFKDFGVSDKSSMFAHIMNVDRHD
HYLDACAPGGKACHIAEVLMPGEGQVDAIDHDKIDLINFNKIKRLNLIKAFQHDA
TKPYDKTYDKILVDAPCSGLGVMRHKEPKYTSQKHIESLVELQLEILENVKNVKI
GGEIIVCTCTIEQLENVIYTFKKNKNFEPEFOHPTIGELVKTQLQIMPQDFNSDG
FITIKKRKN"
4543..5637
/gene="SA1061"
4543..5637
/gene="SA1061"
/note="ORFID:SA1061"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42313.1"
/db_xref="GI:13701018"
/translation="MITAEKKKXKNKTLPNFDKQSYSLRFDENONMLVGGQOKFRK
QIFEWLYQKRVDSIDEMTNLSKDLROLLDNFVTTLTVVVKQESKDGIKFLFELD
GYTIEVLMRHDIYNGVCVTVQVCGICGTFCASGLGKRLNLEAGEIYSOVLTVOKA
LDATERYSVQIVMGIGEPENVDENMDFLRVNDNDSLNGARHITVSTSGIPIRY
DFADEDIQINFAVSLHAKDEVSRLMPINRAYNVKLEIAIQYQEKINRRVTEYG
LFGVNDQLEHARELAHLIRKLNCHLVNLPVNVHPERVNYVKTAKNDIFKFEKELRLG
INATIRREGQSIDAACQLRAKERQVETR"
5644..6387
/gene="SA1062"
5644..6387
/gene="SA1062"
/note="ORFID:SA1062"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42314.1"
/db_xref="GI:13701019"
/translation="MLEAQFFTTQGHDRKDNEDAGGIYNOTNOQLLVLCDSMGHGKA
GEVASKFVTELSRFEAENLIEHQENLNNIKDINFQLYHQAENAYKSGMTT

```

gene	CVCALPEKSVIVANVGDSRAVYINRSRIEQITSDHSFYNNHLVLCQTTPBEAFTHPQ RNIITKMGDKRSPDLFIKRLANFYDYLILNSDGLTDYVKONEIKRLLVKEGTIEH GDOIMOLALDNHSHKDNVTFILAAITBGKV"	
	6384..8378	/gene="SA1063"
CDS	6384..8378 /gene="SA1063" /note="ORFID:SA1063" /codon_start=1 /transl_table=11 /product="protein kinase" /protein_id="BAB42315.1" /db_xref="GI:13701020" /translation="MIGKIINERYKIVDKLGGGMSTVYLAEDTILNIKVAIKAFIFP PREKEETKRFREVENSSOLSHQIVMSIDVDEDDCYLVMEVIEGPTLSEVIESH GPLSDVIAINFTQIILDGIKHAHDMRIVRDIKQNILIDSNTKLIKIPDFGIARLSE TSLQITNEHVLGTQVSPSPQAGEATDECDDIYSIGIVLYEMLVGEPPFNETAVSTA IKHLODSVNTTIDVRKDIPQSLNSVILRATEKDKANRKYITIOEMKDDLSVLHENRA NEDYELDKKRTIAVPLKEDLAKHISHKSNQPKRETTQVPIVNGPAHQHQQFQKPEG TVYPFKKSTKRIKVLISLIPSLMIALYSFVAMAFNGKYEETPDVIGSVKEABQ IFNKNNILKGIKRSYSYDKYPENETIKTNTGTERVERGSDVDVVISGPEKVPWPV IGLPEKALOKLSLGLKDVTIEKVYNNQPKGYIANQSVTANTBIAIHDSNIKLYES LGIQVYVEDPEKSPSKAKALEKGFVESKEEYSDDIDEGDVISQSPKGSVDGEG STISFVSKGKSDSDVKTTTIESVDVPTGKNDKSKQKVYIKDKDNDGSTGKSGPD ITSQORIDPLRIEKGKTASYIVKVDGKTVAEKEVSYDDI" 8606..9481 /gene="SA1064" 8606..9481 /gene="SA1064" /note="ORFID:SA1064" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAB42316.1" /db_xref="GI:13701021" /translation="MATGRIVKISGVVQVDVNGERFNKPKGLFRKKKFPVYVDIV EFEVONIEGIIHQVEPRKNELRPVPSNIDTIVIMSAPENFSTQLLDRFLVIAHS YOLNARVLVKKDKPTIEQFENELLIKVIENGTETFGNDDDRKKIIVBAPWAGLI VLSGQSGVSKTFLNHYRPELNEIDNISLNRKHTTRHVELFERONGYIADTPQF SALDFDHIDKDEIKDYFLENLRYGETCKFRNCNHKEPNCNVKHKQLEIGNIAQRYDH YLOLFNEISNRKY" 9482..10126 /gene="cfxE" 9482..10126 /gene="cfxE" /note="ORFID:SA1065 ribulose-5-phosphate 3-epimerase homolog" /codon_start=1 /transl_table=11 /protein_id="BAB42317.1" /db_xref="GI:13701022" /translation="MTKLYPSLLSYDFLDLQHELKRLERAGVDGVHFDVMDGQFVPNI SIGLPILDVAKEGCTTLPIDVHLMENPEKYIASPBAGADMISIHVESTPHIHRALQM IKHLDKAGVWVNPGRPISQIEPILDIVYLVNVPNGSQSFDIDOCVEKIALGNA IKMERQLNFDIEVDGGVNTDAKVCVENGATMLVTGSFFFKEDYKKVYQALQK"	
gene	3.07e-127 Length: 303750 2274.00 Matches: 455 Score: 100.00% Percent Similarity: 100.00% Best Local Similarity: 100.00% Query Match: 100.00% DB: 1	
CDS	US-09-943-108a-2 (1-455) x AP003133 (1-303750) Alignment Scores: Pred. No.: 2274.00 Score: 455 Percent Similarity: 100.00% Best Local Similarity: 100.00% Query Match: 100.00% DB: 1	
gene	1 MetalaphetCiuLysSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20       	
	Db 26306 ATGGCAATTTGAAGGCTTATCAAGCGCTCTCAAGCGACGATGCAAAAAAATGCGTGGTAAG 26365	
gene	21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40       	
	Db 26366 GGTAACCTTACTGAAGCGTGATTAAGATAATGATCGTGAAGTAAGATTAGCGTTATTT 26425	

QY	41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60       	
	Db 26426 GAGCGTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAAACAGATCAGAACGGCA 26485	
QY	61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80       	
	Db 26486 TTAGGTTCCGATGAATGCAATCATTAACACGAGGCAACAGTTATTAAAAATAGTTCAA 26545	
QY	81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100       	
	Db 26546 GATCAATTACGAAGTTGATGGTGGAGAAAATACATGATTAATATGTCCAAATAACCA 26605	
QY	101 ProThrValValMetMetValClyLeuGlnGlyValGlyLysThrThrAlaGlyLys 120       	
	Db 26606 CCTACTGTGTTGATGATGGTGGTTTACAAGGTGCTGGTAAACAACAACCTGCAGGTAA 26665	
QY	121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140       	
	Db 26666 TTAGCATTATTGATCGTAAAAAATACACAAAAAACCTTATGTTAGTTGCAGCAGATATT 26725	
QY	141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160       	
	Db 26726 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGAAACAAATGATATTCCTGTA 26785	
QY	161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180       	
	Db 26786 TACAGTGAAGGAGATCAAGTAAGCCACAAATGTGAACCTATGCAATTAACACATGCT 26845	
QY	181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200       	
	Db 26846 AAGAAGAACATTTAGACTTTGTTAATCATTATACAGCAGGTGCGATTACACATCGATGAA 26905	
QY	201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220       	
	Db 26906 GCATTGATGAACGAATTAAGAAGAGTAAAAAGAAATTCCTAAACCAACAGCAATATGTTA 26965	
QY	221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240       	
	Db 26966 GTTGTCGATCAATGACGGGTCAAGATGCTGTCAATGTTGCAGATCTTTTGACGATCAA 27025	
QY	241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260       	
	Db 27026 CTTGATGTCACAGGTGTTACCTTAACATAAATTAGATGGTGATACACGTGGTGGCAGGT 27085	
QY	261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280       	
	Db 27086 TTATCTATTTCGTTGCGTGACACAAAAACCAATTAATTTGTTGGTATGAGTGAAGATTA 27145	
QY	281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300       	
	Db 27146 GATGGTTAGAGCTATTCCTGCTGAACGTATGGCATCAGCTATTTTAGGTATGGGTGAT 27205	
QY	301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320       	
	Db 27206 GTGTGAAGTTTAATTTGAAAAACGCAACAGATGTGATCAAGAAAAAGCAAAAGATTTA 27265	
QY	321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340       	
	Db 27266 GAGAAAAAGATGCGTGAGTCATCGCTTTTACTTTAGATGATGATTTTAGAACAACTTGTATCAG 27325	
QY	341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360       	
	Db 27326 GTGAAAAATCTAGGACCACTGGATGATATTATGAAAAATGATTCAGAGGTATGAAATAAATG 27385	
QY	361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380       	
	Db 27386 AAAGGGGTAGATAAGCTTAATATGAGTGAAGCAAAATTTGATCATATTAAAGCGATTATC 27445	
QY	381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400       	
	Db 27446 CAGTCAATGACCGCGCTGAAGAAACAATCCAGACACATGAATGATGATCAGCGTAAAAAG 27505	

QY 401 ArgileAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetIysGlnPhe 420  
 |||||  
 Db 27506 CGTATGCTAAAGGCTGTGCTGTCATTACAGAAGTCATCGTTGATGAACAATTT 27565  
 |||||  
 QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440  
 |||||  
 Db 27566 AACGATATGAAGAAATGATGAACAATCTACTGGTGGCGTAAAGTAAAGGTAA 27625  
 |||||  
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyLysMetAsnLeuProPhe 455  
 |||||  
 Db 27626 CGCAATCAATGCAAAATATGTTAAAGGTATGAATTAACGTTT 27670  
 |||||  
 RESULT 3  
 AP003361  
 LOCUS  
 DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete  
 sequence, section 4/9.  
 ACCESSION AP003361 BA000017  
 VERSION  
 KEYWORDS AP003361.2 GI:14246761  
 SOURCE  
 Staphylococcus aureus subsp. aureus Mu50 (sub\_species:aureus Mu50,  
 strain:Mu50) DNA.  
 ORGANISM  
 Staphylococcus aureus subsp. aureus Mu50  
 Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 REFERENCE  
 1 Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,  
 Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,  
 Matsunari,H., Maruyama,A., Murakami,H., Hosoyama,A.,  
 Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,  
 Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,  
 Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,  
 Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.  
 Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus  
 Lancet 357 (9264), 1225-1240 (2001)  
 21311952  
 2 (bases 1 to 347235)  
 Ohta,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College  
 of Medical Technology and Nursing, Department of Medical  
 Technology; 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan  
 (E-mail:tohata@akura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,  
 Fax:81-298-53-3454)  
 COMMENT On May 29, 2001 this sequence version replaced gi:13875505.  
 FEATURES  
 source  
 1. .347235  
 /organism="Staphylococcus aureus subsp. aureus Mu50"  
 /strain="Mu50"  
 /sub\_species="aureus Mu50"  
 /db\_xref="taxon:158878"  
 35..1021  
 /gene="SAV0992"  
 35..1021  
 /gene="SAV0992"  
 /codon\_start=1  
 /transl\_table=11  
 /product="oligopeptide ABC transporter ATP-binding protein  
 homologue"  
 /protein\_id="BAB57154.1"  
 /db\_xref="GI:14246762"  
 /translation="MNVLVLEKDLWLSKINNEWLVNENIPELSKGVGLVIGES  
 GCKSLSKSIKLLPEKISLSSGVIFDGKIDTANEKQLLDIRNDAMTFQPM  
 TALNPVFTKNOLVESYKSHKISKKEANNLAKDLKVGARDEILNSYPHQLSG  
 MRWRVMAIASCPKLIDAEPTALDVTIOAQLDLKELQKETOAMITHDLS  
 VVAEFCDKVLMYAGQIVIEGGIKTHLPKHPYTKLLISTPKREGQKRTIEGI  
 VPSIQAFHYNCKRFANRCKLIDCNQSPKMHVCEIVIVRCHLYKNEYKEI"  
 1024..2004  
 /gene="appf"  
 1024..2004  
 /gene="appf"  
 /note="SAV0993"  
 /codon\_start=1  
 gene  
 CDS  
 /transl\_table=11  
 /product="oligopeptide transport system ATP-binding  
 protein Appf homologue"  
 /protein\_id="BAB57155.1"  
 /db\_xref="GI:14246763"  
 /translation="MENILEVNOIKKKYIKTGILQKQYVKAVDVVSIRKKGTG  
 LVGSGCKSLGKVIRLEDATSGSIIVNGEDITRLGCKLKRKQQTOMFQPIA  
 SLNPMQMGVDLIISPIILNKKLPKEIKKEVYLILKCVGSEDAIKYIAHFTSGGQR  
 RVGIARALAPSLIVADPEVSALDYSVQSNLLKDLQEFKLSYLETAHDLVVVK  
 HSDVIGVMILGHIVEIASKEIYENPKHPYTKALISSIPOIDKHNHNRILLKGELPS  
 PSNPQCGPFHTRCPIAKCKKENIPQLKDIGNEHQVACFYVDKVGSLNA"  
 1997..2959  
 /gene="oppB"  
 1997..2959  
 /gene="oppB"  
 /note="SAV0994"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable oligopeptide transport system permease  
 protein"  
 /protein\_id="BAB57156.1"  
 /db\_xref="GI:14246764"  
 /translation="MLKILKRLGLMIPLLIISIVFSLAIQDPFSDQLQNGIK  
 ORALNAQREKLNDISPHOYIRWNHVIHGLDGSIKYKRPVIDVTEIRIPNTILG  
 AMSLITVLIISFALGITSGRYSYSLDVTVOIFNYMLAIPSTAGVFAIFESFELQ  
 WFPQGSVDINLKEGTEFYMSKIIYHTFLPALTGLLSTAGYIQYLRNDIENSKDY  
 VIFARSKLSMKNKILNKHILRNSLIPITLITFLGADIVSLGGAVITETIFSYNGIKLF  
 LBSVIGQDYPPLMALTFFSLFGLLGNLISDITYGIDPIRNS"  
 2971..3852  
 /gene="SAV0995"  
 2971..3852  
 /gene="SAV0995"  
 /note="similar to oligopeptide transport system permease  
 protein Oppc"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB57157.1"  
 /db\_xref="GI:14246765"  
 /translation="MQNKSFPKIALSRFINHKIAMLSIIFLLIITVSIAPLIAP  
 FPNQODLNKIGEMTAQNLITGSGDNFSLIYAGRIISLIGTSTIGMLLIGIT  
 VGVISYFGIVDTLLMRITETFVMLFPFLIFAIIVLNAALGDKIKNPYGSAILIVLVII  
 ALSWGIAIRLVKGVLOEKENEYFLAAKSIGTPTYKILKLLPNILSVIVQATILF  
 GIMVVEGSLFGLFGISKAIPSGNMLSDAQEGDVISGKPIWMPAINMIFILSI  
 NFGEGKLDAFNPRGR"  
 complement(3894..4883)  
 /gene="trpS"  
 complement(3894..4883)  
 /gene="trpS"  
 /note="SAV0996"  
 /codon\_start=1  
 /transl\_table=11  
 /product="tryptophanyl-tRNA synthetase"  
 /protein\_id="BAB57158.1"  
 /db\_xref="GI:14246766"  
 /translation="METLFSGIQPSGIPITGNVIGALFKQFVDVNDYDVCIVDQHA  
 ITWPDRLKROKQROLAAIYLAGSDIDPKATLFIQSEVPAHQAGWMLTIAASGEL  
 ERTQYKDKAKAVEGPAGLLIYPPMAADIVLNTNIVPVGDDQKHIELTRNLVD  
 RFSNRVNDLVKPEIRMPKVGVRVNSLIQDPTKMSKSDDNKAFISLLDEPNVAAKKI  
 KSAVTDSDGIIFKFDKDNKPGTINLISIYAGLDMPIKDIEAKYEGEGYKFGKDLAEI  
 VKAFIVFQEKVESFYNSDKLDDILDQGRDKAHKVSFKTVKKMEKAMGLGRK"  
 5177..5572  
 /gene="SAV0997"  
 5177..5572  
 /gene="SAV0997"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB57159.1"  
 /db\_xref="GI:14246767"  
 /translation="MVLFTSPSCTSCRKAKAWLQEHDPYTERNIFSEHLTIDEIKQ  
 ILKMTEDGDEIISTRSKTYOKLNDVDSILPLQDLYSLIQDNPOLLRRPILDNKRLQ  
 VGNEDIRFLPRKVRTFQLQEAQRWVD"

gene

```
5943...6662
/ gene="SAV0998"
CDS
5943...6662
/ gene="SAV0998"
/ note="similar to negative regulator of genetic competence
      MecA"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57160.1"
/db_xref="GI:14246768"
```

```
/translation="MRIERVDVDTTVKLFITVSDIEARGPSREDLWNRKRGEFFWSM
MDEINEEDFVFGPLWIQVHAFKGEVETISKNNEDMMNSDDDATDQDFDQVQL
LAQLEGEQLLELFQRTKEKEAQRKRSKRNKRTIIVKENDLEDVINTAYHS
NPITFERDILLYXVDGTYYAYVHFDVSHVDQEVINDYSQLLEAFYPTDRTEVYLNDYA
KTIKSHNVTAQVRRYPPEITE"
6783...7769
/ gene="SAV0999"
CDS
6783...7769
/ gene="SAV0999"
/ note="similar to transcription factor"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57161.1"
/db_xref="GI:14246769"
```

gene

```
/translation="MLVALNEEKERVLTATTALRKTQYFCVCGKQVILKRGKLVISHP
AKHLAQKCFNNETIKHYKSKLILAQMIQOQCKVIEPPELKEIKQIPDLINKYV
TELQYSPYKQILQRTBGLKMGYKYSWLLNDVYCHNKVFNHPSLFINITRKL
HTFNLEKQIMFQQIQLGGHKYVAERKNAKIIELFNEAPCDIHAVYKLSKFAINQY
IKYCKWNSVLEPTELSAMYQLQLTQDEVVYNYGYIFPEIYIENHPIEQLQVDWLK
NGKSLVDNLNFTFLKKFIVALESKTAIEKLIINYNINICSDRGNDVQILF"
7817...9625
/ gene="SAV1000"
CDS
7817...9625
/ gene="SAV1000"
/codon_start=1
/transl_table=11
/product="thimet oligopeptidase homologue"
/protein_id="BAB57162.1"
/db_xref="GI:14246770"
```

gene

```
/translation="MSQQLSREEQERKYPETWDLTTFIKDDAEAPAEKVEYNELK
BEQFGHIGDSASTLYNALEEDTIGTKLVYVYAHKQDQDQDNDKRYTGMESRAHQ
LIIFESAWGSLVPEILQIDEDKIQSFVNSYDKLQKFAFDLKLINKEKPHILDAETK
LITEAQAALSTPSNVGMFSNADLVFEDAIDKGNAPLFOGTFIKYLESDDRKLRIS
AFPNRYAGAHNNTLGATLAGEVKVFNARTNYKTAREKALSNHNHPENVYDNLV
KTVKHYLPPLHRTTELKELGLDOLKMYDLYTPLIDKFEMPYEAKEMWALKALEP
XGEVLYNVREGLNRRWDVYENKRGSGGSGSAHLNPFILLNSNTISDLTVH
EFGHSASHYSRKFPQSNSDYITFVAEVASTCNEALLSDIMDKHLDDEKRLLLNQE
LERFRATLFRQTFMFAEFERKIHAIIEAGEPELTPTRMNEEYAKLNKLYFGDSVETDEI
SKWESRIPHFYMYVYOATGYSAQSLSHQILTEGKPAVDRYINEFLKKGSSNYPI
EILKNAGVDYMTPEPTEQACEVFEQKLNAFEKLMKA"
complement(10085...10639)
/ gene="yjbM"
complement(10085...10639)
/ gene="yjbM"
/ note="SAV1001"
/codon_start=1
/transl_table=11
/product="GTP pyrophosphokinase"
/protein_id="BAB57163.1"
/db_xref="GI:14246771"
```

gene

```
/translation="MKQLSYKVVSKETIYTYTCNBIIIPKRDIIIPESMICDQIAG
IDLEVFKDDLOKSLTESKLDLHIAREMEIEQAPSLVFFSEVDVHEGLKVBGLYPVH
ITYTYINELMGKPTKKNLPKLETYIQOQLVTEBELLTIYEWPEKLNKELKLAIQ
QXIEKLYPDGDFWKSMPKIKSK"
complement(10575...10901)
/ gene="SAV1002"
```

gene

```
/translation="MKQLSYKVVSKETIYTYTCNBIIIPKRDIIIPESMICDQIAG
IDLEVFKDDLOKSLTESKLDLHIAREMEIEQAPSLVFFSEVDVHEGLKVBGLYPVH
ITYTYINELMGKPTKKNLPKLETYIQOQLVTEBELLTIYEWPEKLNKELKLAIQ
QXIEKLYPDGDFWKSMPKIKSK"
complement(10575...10901)
/ gene="SAV1002"
```

gene

```
/translation="MKQLSYKVVSKETIYTYTCNBIIIPKRDIIIPESMICDQIAG
IDLEVFKDDLOKSLTESKLDLHIAREMEIEQAPSLVFFSEVDVHEGLKVBGLYPVH
ITYTYINELMGKPTKKNLPKLETYIQOQLVTEBELLTIYEWPEKLNKELKLAIQ
QXIEKLYPDGDFWKSMPKIKSK"
complement(10575...10901)
/ gene="SAV1002"
```

gene

```
/translation="MKQLSYKVVSKETIYTYTCNBIIIPKRDIIIPESMICDQIAG
IDLEVFKDDLOKSLTESKLDLHIAREMEIEQAPSLVFFSEVDVHEGLKVBGLYPVH
ITYTYINELMGKPTKKNLPKLETYIQOQLVTEBELLTIYEWPEKLNKELKLAIQ
QXIEKLYPDGDFWKSMPKIKSK"
complement(10575...10901)
/ gene="SAV1002"
```

Alignment Scores:

Pred. No.: 3 57e-127 Length: 347235

Score: 2274.00 Matches: 455

Percent Similarity: 100.00% Conservative: 0

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
```

US-09-943-108a-2 (1-455) x AP003361 (1-347235)

QY 1 MetaAlaPheGluGlyLeuSerGluArgGluGlnAlaThrMetGlnLysMetArgGlyLys 20  
|||||  
Db 256730 ATGCGATTGAAGGCTTATCAGACGGCTCGAAGCGCGCATGCAAAAAATCGGTGGTAA 256789  
|||||

QY 21 GlyLysLeuThrGluAlaAspLysLysLysMetArgGluValArgLeuAlaLeuPhe 40  
|||||  
Db 256790 GGTAACTTACTCAAGCTGATATAAGATAATGATCGTCAAGTAGATTAGCGTTATTT 256849  
|||||

QY 41 GluAlaAspValAsnPheLysValValLysGluPheLysThrValSerGluArgAla 60  
|||||  
Db 256850 GAGCTCAGCTAACTTTAAAGTGTGTAAGAAATTTATTAAACACAGTATCAGACGCGCA 256909  
|||||

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysValGln 80  
|||||  
Db 256910 TTAGTTCCGATGTAATGCAATCATTTACACACGAGGCAACAAGTTATTAAATAGTTCAA 256969  
|||||

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
|||||  
Db 256970 GATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTAAATATGTCAAATAAACCA 257029  
|||||

QY 101 ProThrValValMetMetValGlyLeuGluGluAlaGlyLysThrThrAlaGlyLys 120  
|||||  
Db 257030 CCTACTGTTGTTATGATGGTGGTTTACAGGTGCTGGTAAACACCACTGCAGGTAAA 257089  
|||||

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
|||||  
Db 257090 TTACGATTAATGATCGTAAAAAATACAAAAAACCTATGTTAGTTGCAGCAGATATT 257149  
|||||

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnLysLysProVal 160  
|||||  
Db 257150 TATCGTCCAGCAGCAGATAAATCAATTACAAACAGTAGGGAACAATGATATTCCTGTA 257209  
|||||

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
|||||  
Db 257210 TACAGTGAAGCAGANTCAAGTAAGCCACACAAATTTGTAATGCAATTAACACATGCT 257269  
|||||

QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
|||||  
Db 257270 AAAGAACAACATTTAGACATTTGTAATCATTCATACACAGAGTCGATTACACATCGATGAA 257329  
|||||

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
|||||  
Db 257330 GCATTGATGAACGAATTAAGAAAGTAAAGAAATGCTAAACCAACGAAATATTGTTA 257389  
|||||

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
|||||  
Db 257390 GTTTCGATTCAATGACGGTCAAGATGCTCAATGTTGCAGATCTTTTGACGATCAA 257449  
|||||

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
|||||  
Db 257450 CTTGATGTCACAGGTGTTACCTTAACATAAATTAGATGGTGATACACAGTGGTGGTCAAGCT 257509  
|||||

QY 261 LeuSerIleArgSerValThrGluLysProIleLysPheValGlyMetSerGluLysLeu 280  
|||||  
Db 257510 TTATCTATTCGTTTCGGTGACACAAAACCAATTAATTTGTTGGTATGAGTGAAAGTTA 257569  
|||||

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
|||||  
Db 257570 GATGGTTTAGAGCTATTCCTGACAGCTATGGCATCAGTATTTTAGGTATGGGTATG 257629  
|||||

QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
|||||  
Db 257630 GTGTTAAGTTTAATTCAAAAAGCGCAACAAGATGCGATCAAGAAAAAGCAAAAGATTTA 257689  
|||||

QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340  
|||||  
Db 257690 GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTGTAGAACAACTTTGAG 257749  
|||||

QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 257750 GTGAAAATCTAGGACCATCGATATATGAAAATGATTCACGATGATGAAAATG 257809  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGlyLysGlnIleAspHisIleLysAlaIleIle 380  
 Db 257810 AAAGGCTAGTAACTTAATAGTGAAGAAGCAAAATGATCATATTAACCGATATC 257869  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 257870 CAGTCAATGACCGCGGTGAAGAACAATCCAGACACATGATGATCAACGTAAGAAG 257929  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
 Db 257930 CGTATGCTAAAGGCTGCTGCTCATCAAGAAGTCATCGTTGATGAACAATTT 257989  
 QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440  
 Db 257990 AACGATATGAAGAAATGATGAACAATCTACTGTCGCGTAAAGTAAAGTAAGGTAAA 258049  
 QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMetAsnLeuProPhe 455  
 Db 258050 CGCAATCAATGCAAAATATGTTAAAGGTATGATTAACCGTTT 258094  
 RESULT 4  
 AP004826 295350 bp DNA linear BCT 02-JUL-2002  
 LOCUS Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,  
 DEFINITION strain:MW2, section 5/10.  
 ACCESSION AP004826 BA000033  
 VERSION AP004826.1 GI:21204263  
 KEYWORDS  
 SOURCE Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.  
 ORGANISM Staphylococcus aureus subsp. aureus MW2  
 Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 REFERENCE 1  
 AUTHORS Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,  
 Nagai,Y., Iwana,N., Asano,K., Naimi,T., Kuroda,H., Cui,L.,  
 Yamamoto,K. and Hiramatsu,K.  
 TITLE Genome and virulence determinants of high virulence  
 JOURNAL Lancet 359 (9320), 1819-1827 (2002)  
 MEDLINE 22040717  
 PUBMED 1204378  
 REFERENCE 2 (bases 1 to 295350)  
 AUTHORS Aoki,K., Oguchi,A., Nagai,Y., Asano,K., Iwana,N., Baba,T.,  
 Kuroda,M., Hiramatsu,K. and Kikuchi,H.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAR-2002) Akio Oguchi, National Institute of  
 Technology and Evaluation, Biotechnology Center; 2Chome 49-10  
 Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan  
 (E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/,  
 Tel:81-3-3481-8423, Fax:81-3-3481-8424)  
 FEATURES  
 Location/Qualifiers  
 1. 295350  
 /organism="Staphylococcus aureus subsp. aureus MW2"  
 /strain="MW2"  
 /db\_xref="taxon:196620"  
 301..1254  
 /gene="MW1096"  
 301..1254  
 /gene="MW1096"  
 /note="ORFID:MW1096"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB94961.1"  
 /db\_xref="GI:21204264"  
 /translation="MKKTGLCLLLMLVAVGSGFNHLSKSSKEESKQFTVKES  
 EEKQPDLEKYEIEKMKGIKDAISLDKLDPLMTKSFNSKGIQGWKIDKELMGKV  
 ELADYFRDQSGSIKDVDAFFKGGIKRKVIETHDDVKQVDYNTVPDQKIGNS  
 NTPVFVAEIMTKYKDGKLVVAVSGSVYIHKDDAIKYDDYSKLILSLITLIDHPKP

VPYSVAQIKSRGVPLTSTVSMTHGSKDTEVLPAALAYTFSTPKANEDSKNPDPKVLN  
 LVHMFPLAASDFGNHAFVVLISKYIKKEYSTETASDDSLK"  
 complement(1417..1695)  
 /gene="MW1097"  
 complement(1417..1695)  
 /gene="MW1097"  
 /note="ORFID:MW1097"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB94962.1"  
 /db\_xref="GI:21204265"  
 /translation="MPELKVGIIKREANVIISNTEGKEMQVKKVYIILLALFLGSFG  
 IHKFAKGMGIIILHIFCWTGIPHILAIISAVITVFKPADEQGNVTL"  
 1919..2407  
 /gene="MW1098"  
 1919..2407  
 /gene="MW1098"  
 /note="ORFID:MW1098"  
 hypothetical protein, similar to polypeptide deformylase"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB94963.1"  
 /db\_xref="GI:21204266"  
 /translation="MAIKLYPASHPIELTKAQAIVIKFDDSLKLLQDLEDTHYAGEA  
 AGLCAPQINSLQVAIIDMEMEGILLQVNPKIISQSNETIIDLEGSITLFDVYGEYTR  
 SKVIVSYDYNKVELTAHEDVARMILHIIDQMGPIPTERADRILTDKEVEAYFI  
 ND"  
 2400..3335  
 /gene="MW1099"  
 2400..3335  
 /gene="MW1099"  
 /note="ORFID:MW1099"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methionyl-tRNA formyltransferase"  
 /protein\_id="BAB94964.1"  
 /db\_xref="GI:21204267"  
 /translation="MTKILFWGPPDSTVLEMLIAEHVIAVVTOPDRVGRKRVMT  
 PPVKKVAMKYDLFVYQFEKLSGELEQLLDVLIYVTAAGQLLPSLILALPKLG  
 AINVASLLPKYRGGAFTHOAIIDGQETGITIMYVKKLDAGNIISQAIIEENDN  
 VTGMDKLSVLGADLLKETLPSIIIEGTNESVPQDDQATFASNIRREDERINWPGR  
 QVFNQIRGLSPWPVAVTMDDTNLKIYDAELVETKINIEPCTIETTKAIIAINDN  
 EAVAIDMQLAGKXKMLAANYLSGAQNTLVGKKLI"  
 3332..4639  
 /gene="MW1100"  
 3332..4639  
 /gene="MW1100"  
 /note="ORFID:MW1100"  
 hypothetical protein, similar to RNA-binding Sun protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB94965.1"  
 /db\_xref="GI:21204268"  
 /translation="MIENVRSLAFDTIODILNEGAYSNLRINEVLSLENLAMDKALF  
 TRIVGVTKRKETLDLYLKPVYTKIKAWQOLLWMSLYQVYVLDKVPNALINEAVE  
 IAKRGCGYHNGVNGVILRTMRSDLDFNEIADPKRMAIEISMPKWIIDHWATHYG  
 LKRTILOSEFETTTSTVRANLTRASLDIIIEKLQDEYDEKDHDLPCYLIHGQGP  
 ITHSRFKDGFVSQDSSVFVHIMNVDRHDHVLDCASAPGGKACHIAEVLHPGGQV  
 DASDIHDHIDLINFINIKLRITNKAFOHDKTPYDKTYDKILVDAPCSGLGVNRHK  
 PEIKYTSQKHIESLELOLEILENVKNVKNKIGGELIYSTCTIQLENENVIYTLKN  
 NKNEFEFPQHPITGELVLTQLINFDQNSDGFPTTKIKKRDN"  
 4642..5736  
 /gene="MW1101"  
 4642..5736  
 /gene="MW1101"  
 /note="ORFID:MW1101"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB94966.1"  
 /db\_xref="GI:21204269"  
 /translation="NITAEEKKKKKFELPNFDKQSIYSIRFDEMQLNVLVEQGOQKPRAK







QY	381	GlnSerMetThrProAlaGluArgAsnValSerProAspThrLeuAsnValSerArgLysLys	400
Db	526	RAAGTCAATGACAAAAATCAAAAGACAAATCCAGATATCATCAATGCAAGTGAAGAAAA	467
QY	401	ArgIleAlaLysGlySerGluArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe	420
Db	466	CGGATTCGACGTGGTACGCGACTCCGGTTCAGAGAATTACCGTCTCTCTTAACCAATT	407
QY	421	AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys	440
Db	406	GGTGAAATGAAAAAATGATGAAGCAATCACTGGTGGCGCAAGGCAAAAGGCAAA	347
QY	441	ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe	455
Db	346	-----AATCATTCGCGAATTCAAAATGCCAATT	317
RESULT	6	AL591981	347050 bp DNA linear BCT 06-JUN-2002
LOCUS	AL591981/c	AL591981	Listeria monocytogenes strain BGD, complete genome, segment 9/12.
DEFINITION	AL591981	AL591981	
ACCESSION	AL591981	AL591981	
VERSION	AL591981.1	GI:16411141	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A.,			
Bequero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T.,			
Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P.,			
Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,			
Dussurget, O., Entian, K. D., Psihi, H., Portillo, F. G., Garrido, P.,			
Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J.,			
Jackson, D., Jones, L. M., Kaerst, U., Kref, J., Kuhn, M., Kunst, F.,			
Kuraokat, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E.,			
Nedjari, H., Nordieck, G., Novella, S., de Pablo, B., Perez-Diaz, J. C.,			
Purcell, R., Rammel, B., Rose, M., Schluter, T., Simoes, N.,			
Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J., and Cossart, P.,			
Comparative genomics of <i>Listeria</i> species			
Science 294 (5543), 849-852 (2001)			
21537279			
PUBMED	11679669		
REFERENCE	2 (bases 1 to 347050)		
AUTHORS	Glaser, P., Frangeul, L. and Rusniok, C.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des		
	Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris		
	Cedex 15, FRANCE		
COMMENT	E-mail: pglaser@pasteur.fr		
FEATURES	Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.		
source	1. 347050		
	Location/Qualifiers		
	/organism="Listeria monocytogenes"		
	/strain="EGD-e"		
	/db_xref="taxon:1639"		
gene	complement(95..841)		
	/gene="lmo1688"		
CDS	complement(95..841)		
	/note="similar to glucose 1-dehydrogenase"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAC9766.1"		
	/db_xref="GI:16411142"		
	/db_xref="SPTREMBL:O876J6"		
	/translation="MNKVALDVTGSSRGGRGRIATALAKGYDIAVFNRRKKAEVQ		
	QETQGLRKQVIFKAVDYKAEKVLKFAVDDEERGLDIFINNAAGVLRPLMLELRS		
	HWDVTKNTAKALLFAQCAKALMOPHOSGKIIISIGSIRGLVNTYTVGVSKAAVE		
	SLTRYLAVLAPFGIAYNAVSSGLITETDALNHFFNRELLKDAVSKTPAGRMIPNDL		
	VNKVFLFLSEKADMRIGQTLIVDGGRTLLV"		
gene	complement(join(845..1942,1948..1953))		
	/gene="lmo1689"		
CDS	complement(845..1942)		

/note="similar to hypothetical proteins"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC99771.1"

/db\_xref="GI:16411147"

/db\_xref="SPTREMBL:Q8Y6J1"

/translation="MKITISVQOKNKRNYIFIDERYNSVDEEVLARYOLMKGVKLT  
TRADIERIKQADMYRKLGNKALNFSLHRVSEKRIIDYLRKQEMPEFAIDELKLAN  
MDYINDEFEALTKYQIKTKLGPRTIERELVEKGLTREIISRVIRESDBAQLENA  
TKQAKIMKRNKNSAKMLQOKIITLDLQKFTSELAKAAATSELDVADAEEILQ  
KOIEKTMKRRKYPSTAKOKITSLMKQGFSDYTIQSYLTENETSFEEH"

/complement(3872..3877)

/gene="lmo1692"

/complement(3872..3877)

/gene="lmo1692"

/complement(4680..4685)

/gene="lmo1693"

/join(4757..5673,5681..5708)

/gene="lmo1694"

4757..4762

/gene="lmo1694"

4771..5673

/gene="lmo1694"

/note="similar to CDP-abequose synthase"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC99772.1"

/db\_xref="GI:16411148"

/db\_xref="SPTREMBL:Q8Y6J0"

/translation="MHILITGATGFTGDLHVELEKSDHELYILTROKLNKRNHYI  
EWNLNDLSPEDVPCINLAGLMDKWTYDRKVTVNSRIEATSAIISVYKMK  
KSKPLWNAIGACVYSSKSTIYLDTENYADNFKGVYWEKTAASADLGIYV  
VYARGVLGTNGSSPFVFEKLEPQIYTGGRFGNGRQWFSWIHVDVVAALLIPDHEQ  
INGVNFAPHPVDEKFAERKQKMKPKYKTPVKKIIFILGERAMTILDSQRAYP  
EKLMSNHEFFEFYEQALDDLLD"

terminator

5681..5708

/gene="lmo1694"

/complement(join(5682..8291,8299..8304))

/gene="lmo1695"

/complement(5682..5709)

/gene="lmo1695"

/complement(5694..8291)

/gene="lmo1695"

/note="similar to putative membrane proteins"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC99773.1"

/db\_xref="GI:16411149"

/db\_xref="SPTREMBL:Q8Y6I9"

/translation="MKELMQAVAFQKNSTVVKIVFTFVMAFVIFELINATGIDY  
PSLKNLSQSPELFTMFIVGLIAYTPMLYDYVIVKLLPKGKSPSHVTSAGLNT  
FTNIGFGVIGLSIRASFTGKNASHKEILLALSKIALFLVSGLSYICVLSLATLIP  
GRADHFVYNWLLAGDLYPFIITITKWSKSLFVDLPINKRELTLLIASLLEWGF  
GCFATIGLMEGPDYDFKPLFVIAISVIGIASVPGVGTFDVMILGSLQVSOE  
LALANMLFVIRYFIIPFVWGLFFVQKAGKVNDFLEGLFLFLQKVAERFLIVFY  
GSLGILLSSAPNAIYVPLVYKIMPFNFSTOITIVAFGLLGLLARGTECKTKK  
AXITIVVLCGAIYENTLARYESMKQALFGLIVLCLFLARNEFVREKIVYVTSKVIID  
SIIFIVLCAGIYIGIYNSPNIKHSKEIPDIRIASHELWVGFGVFIYVSVLVIY  
IYLSITKELSGSPFAVREHLKAGNVEVSHTFMRDLKLLFWAAGEVLEFSYRLLA  
DKMVMGEPTGNMOKMAAIEEVMNADRGYRPFVYEVGMTIPYIYHHDGFEDKLG  
EEGFDVQNFQNTMSKKKKGRALMKLEREGYTFEIIPEFPNHDWTYIYHDSWDEL  
PSGIMDFLFINLPEKAKEDGFTFNAGMAPLANGVESKYAFELGRAGIVYRYSOGFY  
GPKGLNFKSKIVYEWQKQFVAFKRSSTAFYMLQIMLVCKRPLANSOVLDPLLE  
EETKPKDSE"

/complement(8299..8304)

/gene="lmo1695"

/complement(8318..9007)

/gene="lmo1696"

/complement(8318..8992)

/gene="lmo1696"

/note="similar to unknown proteins"

Alignment Scores:

Pred. No.: 2,22e-93 Length: 347050

Score: 1717.50 Matches: 331

Percent Similarity: 86.81% Conservative: 64

Best Local Similarity: 72.75% Mismatches: 55

Query Match: 75.53% Indels: 5

DB: 1 Gaps: 1

US-09-943-108a-2 (1-455) x AL591981 (1-347050)

QY 1 MetalPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
Db 122116 ATGGCATTTGAAGGACTAGCTGGAAGACTCCCAAGAACATGAACAAAATTCGGGTA 122057  
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaPhe 40  
Db 122056 GGGAAAGTAACGAAGCTGACGTAAAGAAATGATCGTGAAGTTCGGCTTGCTCTACTA 121997  
QY 41 GluAlaAspValAsnPhelYsValValLysGluPheIleLysThrValSerGluArgAla 60  
Db 121996 GAAGCCGATGTTAACTTTAAAGTCGTTAAACAATTTATTAACACAGTAAGCGAGCT 121937  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
Db 121936 GTCCGCGCGGACGTTATGAAAGCCTAAACCCGCCACACAGTTATCAAAATCGTTCAA 121877  
QY 81 AspGluLeuThrLysLeuMetGlyGluAspThrSerIleAsnMetSerAsnLysPro 100  
Db 121876 GAAGAAGTAACAGCTTAATGGCGGAGAGAAAGAAATCGGAACAGCGGACCGCGG 121817  
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
Db 121816 CCACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 121757  
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
Db 121756 CTTGCTAATTTATGACGAAAATATATACCGTAAACCATTTCTTCACAGAGATATT 121697  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
Db 121696 TATCGACCTGCCGATCAACAATTAAGAACCTTGGCAAGCAATTAAGATATGCGGTA 121637  
QY 161 TyrSerGluGlyAspGlnValLysLysProGlnIleValThrAsnAlaLeuLysHisAla 180  
Db 121636 TTTTCCCTAGGAGACCAAGTAGCCGCGTCAAGTCCCAACCAAGCTATTCTTAAAGCA 121577  
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisLeuAspGlu 200  
Db 121576 AAAGAAGAACAATTTAGATATTGCTATTATGATACACCGCGTCTCTCATATTGACGAA 121517  
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
Db 121516 AGCTTAATGGATGAGTTAAACAGTGAAGAAATTCACAGCCACACCGAATCTTACTT 121457  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
Db 121456 GTAGTCGATTCATACCGCGCAAGTCTGTCAACGTGGCGCAAGCTTCAACGACAA 121397  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260  
Db 121396 TTAGAAATACAGGTGTCGTTATTAACGAATAGATGATGATGATGATGATGATGATG 121337  
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
Db 121336 CTTTCGATTCGTTCCGTCACCTGCTAAACCAATTAATTTATCCGACTCGCGAATAATG 121277  
QY 281 AspGlyLeuGluPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
Db 121276 GAAGCCCTTGAACCTTCATCCAGATCGTATGCTTCCAGAAATTCCTGCGATGGGAT 121217  
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
Db 121216 GTGCTTTCCTGATTGAAAGCGCAACCGATGTAGATACAGAAAAATGAAGCTATG 121157





```

Db 138781 TTAGAAATTACCGCGTGTATTAAACAAATTTAGACGGTGATACACGTTGGTGGCGAGCA 138722
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 138721 CTTTCATCCGTTTCAGTCACAGAAACCAATCAATTCGTTGCTACCGGTGAAAAATG 138662
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 138661 GAAGCAATCAAGAACTCCATCCGATCCGATGCTTCAAGAAATTCGCGCTGGGTGAT 138602
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValaspGlnGluLysAlaLysLeu 320
Db 138601 GTACTTTCTCTATGTAAAGACAACTGATGTAGATGCAGAAATGAAAGCTATG 138542
QY 321 GluLysLysMetArgGluSerSerPheThrLeuaspPheLeuGluGlnLeuaspGln 340
Db 138541 GAACAAAAATGAAGACACACGATGACCTTAGATGACTTCTTGACCAATTTGCAACAA 138482
QY 341 ValLysAsnLeuGlyProLeuaspPheMetLysMetIleProGlyMetAsnLysMet 360
Db 138481 GTAAACAAATGGGACACTAGATGAATCTACTAAAAATGATGCCAGGGCAACAAATG 138422
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
Db 138421 AAAGCCCTCGAACATGCAAGTAGATGATAACAACTCGGTCACTCGAAGCGATAAT 138362
QY 381 GlnSerMetThrProAlaGluArgAsnProaspThrLeuAsnValSerArgLysLys 400
Db 138361 AAATCCATGACCAAAAACGAAAGATAATCCGACATCATCATCGGACGAGAGAAA 138302
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 138301 CGAATTCCTGCGAAGCGACGCCAATTCAGAAATCAATCGCTCTCTAAACAATTT 138242
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
Db 138241 GCTGAATGAAAAAATGATGAGCAATGACTGGTGGAGAAAGGTAAGAAGGTAAA 138182
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 138181 -----AATCCATTGGGCAATTTCAAAATGCCATTT 138152

RESULT 8
AX413016 319630 bp DNA linear PAT 14-JUN-2002
LOCUS AX413016
DEFINITION Sequence 7 from Patent WO0228891.
ACCESSION AX413016
VERSION AX413016.1 GI:21445474
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Parent: WO 0228891-A 7 11-APR-2002;
PASTEUR Institut (FR)
FEATURES
Location/Qualifiers
1. 319630
/organism="Listeria innocua"
/db_xref="taxon:1642"
BASE COUNT 105207 a 55428 c 66726 g 92263 t 6 others
ORIGIN
Alignment Scores:
Pred. No.: 7 12e-93 Length: 319630
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 6 Gaps: 1

```

US-09-943-108a-2 (1-455) x AX413016 (1-319630)

```

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 187924 ATGGCATTTGAGGACTAGCTGGAGACTCCCAAGAAACAATGAACAAATTCGCGTAAA 187983
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 187984 GGGAAAGTAAACGAAGCTGACGCTAAGAAAGAAATGATGCGTGAAGTTCGTTCTTACTA 188043
QY 41 GluAlaaspValasnPheLysValValLysGluPheIleLysThrValSerGluargala 60
Db 188044 GAAGCCGATGTATTAAATCGTTAAACAAATTTAATAACAGTAAAGCAAGTCA 188103
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 188104 GTTGGCGGGACGCTGTATGAAGAAACCTACACCGGTCAACAGTAATTAATAATCCTTCA 188163
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluasnThrSerIleasnMetSerAsnLysPro 100
Db 188164 GAAGAATTAACAAGCCTCATGTGGCGCGGAAGAAAGTAAATCGGAACAGCTGATCGCCG 188223
QY 101 ProThrValValMetMetValGlyLeuGlnGlnAlaGlyLysThrThrThrLagLys 120
Db 188224 CCAACCGTTATTATGATGTTAGTGTGTACAAAGGAGTGGTAAAAACAACCACTTCAGGAAA 188283
QY 121 LeuAlaLeuLeuMetArgLysLysTyraAsnLysProMetLeuValAlaAlaAspIle 140
Db 188284 CTCGCTAATTTATACCTAAAAAATAATATCGTAACCTTTACTAGTCGACAGATATT 188343
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 188344 TACCGACCTGCGACGAATCAACAAATTAGAAACACTTGGCAACAATTAGATATGCCAGTA 188403
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 188404 TTTTCTTAGCGCATCAAGTAAGCCCAAGTAGAAATPCGGGAACAAGCTATPCGCTAAAGCA 188463
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrLagIlyArgLeuHisIleaspGlu 200
Db 188464 AAAGAAGAACAATTTAGATTATATGTCATTATCATGACAGCTGGTCTTCATATCGACGAA 188523
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProasnGluIleMetLeu 220
Db 188524 ACTCTGATGGAGCAATTAACAAAGTGAAGAAATTCGTACGCCAAGTCAAAATTTTACTT 188583
QY 221 ValValaspSerMetThrGlyGlnaspAlaValasnValAlaGluSerPheaspGln 240
Db 188584 GTAGTTGATTCATGACTGGCGAAGCGGTAATGTGGCCCAAGCTTCAACGAACAA 188643
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuaspGlyAspThrArgGlyAlaAla 260
Db 188644 TTAGAAATTCGCGCGTTGATTAAACAAATTTAGACGGTGATACAGTGGTGGGCAGCA 188703
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 188704 CTTTCATCCGTTTCAGTCACAGGAAACCAATCAAAATTCGTTCTACCGGTGAAAAATG 188763
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 188764 GAAGCAATCGAAACGTTCCATCCGATCGATGCTTCAAGAAATTCGCGATGGGTGAT 188823
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValaspGlnGluLysAlaLysLeu 320
Db 188824 GTACTTTCTCTATTGAAAAAGCACAACTGATGTAGATGCAGAAAAAATGAAGCTATG 188883
QY 321 GluLysLysMetArgGluSerSerPheThrLeuaspPheLeuGluGlnLeuaspGln 340
Db 188884 GAACAAAAATGAAGACAAACAGCATGACCTTAGATGACTTCTTGGACCAATTCGAACA 188943
QY 341 ValLysAsnLeuGlyProLeuaspPheMetLysMetIleProGlyMetAsnLysMet 360
Db 188944 GTAAACAAATGGGACCACTAGATGAATCTATAAATATGATGCCAGGGGCAACAAATG 189003

```

```

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189004 AAAGCCTCGACAACTGCAAGTAGATGAACAACACTGGTCACATCGAAGCGTAATT 189063

QY 361 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189064 AAATCCATGACCAAAACGAAAGAGATAATCCGGACATCATCAATCGCAGCAGAAGAAA 189123

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189124 CGAATTCGTGTAAGAGCGGCGCAATTCAGAAATCAATCGCTCTCTAAACAAATTT 189183

QY 421 AsnAspMetLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189184 GCTGAATGAAGAAATGATGACCAATGACTGGTGGAGGAAAGGTAAGAAAGGTAAA 189243

QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMetAsnLeuProPhe 455
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189244 -----AATCCATTCGGCAATTTCAAAATGCCATTT 189273

RESULT 9
AX417046/c AX417046 349980 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 4037 from Patent WO0228891.
ACCESSION AX417046
VERSION AX417046.1 GI:21449656
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
REFERENCE
1
AUTHORS Glaser,P. and Kunst,F.
TITLE Listeria innocua genome and applications
JOURNAL Patent: WO 0228891-A 4037 11-APR-2002;
Pasteur Institut (FR)
FEATURES
source
1..349980
Location/Qualifiers
/organism="Listeria innocua"
/db_xref="taxon:1642"
/note="seq 2038, original length: 3,011,208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"
BASE COUNT 101055 a 72969 c 60688 g 115268 t
ORIGIN

Alignment Scores:
Pred. No.: 7,89e-93 Length: 349980
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 6 Gaps: 1

US-09-943-108a-2 (1-455) x AX417046 (1-349980)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129501 ATGGCATTTGAGAGACTACTCGAGACTCCAGAACATGACAAATTCGGCGGTAAA 129442

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129441 GGGAAAGTAAACGAAGCTGACCTAAAGAAATGATGGGTGAGCTTCGCTTACTA 129382

QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129381 GAAGCGATGTTAACTTAAAGTCGTTAAACATTTATTAAACAGTAAGCGAAGCGTGA 129322

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGln 80
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129321 GTTGGCCGAGCGCTATGAAGAGCCTACACCCGCTCAACAAGTATTAAATGTTCAA 129262

QY 81 AspGluLeuThrLysLeuMetGlyGlyGlnAsnThrSerIleAsnMetSerAsnLysPro 100
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129261 GAAGATTAACAAGCCTCATGTGGCGGCAAGAAATAAATCGGAACAGCTGATCGCCG 129202

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129201 CCACCGCTTATTATGATGTAGTTTACAGGAGCTGGTAAACAACACCTTCAGGAAA 129142

QY 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129141 CTCCTTAATTATAGTAAAAAATAATATCGTAACCTTTACTAGTCGACGAGATATT 129082

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129081 TACGAGCTGCGAGCAATCAACAATAGAACACTTGGCAACAATTAGATATGCCAGTA 129022

QY 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129021 TTTTCTTAGCGCATCAAGTAAGCCAGTAAAGATCGGAAACAAGCTATCGCTAAAGCA 128962

QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128961 AAAGAAGACATTTAGATTATGCTATTATCGATACAGCTGGTGGCTTCATATCGAGAA 128902

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128901 ACTCTGATGCGAGATTAACAAGTAAAGAAATCGTACGCCAACTGAAATTTTACTT 128842

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128841 GTAGTGTATTCAATGACTGGGCAAGACGCAATAATGTGGCCAAAGCTTCAAGCAAA 128782

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128781 TTAGAANTACCGCGCTTGTATTACAAATACCGGTGATACACGTGGTGGGCGACGA 128722

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128721 CTTTCCATCGGTTCACTACAGAGAAACCAATCAATTCGTTCTACCGGTGAAAAAATG 128662

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetClyAsp 300
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128661 GAAGCAATCGAAACGTTCCATCCGATCGTATGCTTCAAGAATTCGCGCATGGGTGAT 128602

QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128601 GTACTTCTCTTATGAAAAACACAACTGATGATCGAGAAAAAATGAAGCTATG 128542

QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128541 GAACAAAAAATGAAGACAAACAGCATGACCTTAGACTTCTTGGACCAATTCGAACAA 128482

QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128481 GTAAACAAATGGGACCCTAGATGACTACTAAAAATGATCGCAGGGGCAACAAATG 128422

QY 361 LysGlyLeuAspLysLysMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128421 AAAGCCTCGACAACTGCAAGTAGATGAACAACACTCGTTCACATCGAAGCGTAATT 128362

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128361 AAATCCATGACCAAAACGAAAGAGATAATCCGGACATCATCAATCGCAGCAGAAGAAA 128302

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128301 CGAATTCGTGTAAGCGGCGCAATTCAGAAATCAATCGCTCTCTAAACAAATTT 128242

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```



Db 128241 GCTGAATGAAAAAATGATGAAGCAATGACTGGTGGAGGAAAGGTAGAAGGTAAA 128182

QY 441 ArgasnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
||||| :|||||

Db 128181 -----ATCCATTCCGCAATTTCAAAATGCCATTT 128152

RESULT 10  
D14356  
LOCUS  
DEFINITION  
D14356 2739 bp DNA linear BCT 03-FEB-1999  
Bacillus subtilis orf1, ffh, rpsp genes for ORF1, ffh and 30S  
ribosomal protein S16, complete cds.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
1 (bases 1 to 2739)

REFERENCE  
AUTHORS  
TITLE  
Honda, K., Nakamura, K., Nishiguchi, M. and Yamane, K.  
Cloning and characterization of a Bacillus subtilis gene encoding a  
homolog of the 54-kilodalton subunit of mammalian signal  
recognition particle and Escherichia coli Ffh  
J. Bacteriol. 175 (15), 4885-4894 (1993)

JOURNAL  
MEDLINE  
REFERENCE  
93328695  
2 (bases 1 to 2739)

AUTHORS  
TITLE  
Yamane, K.  
Direct Submission  
Submitted (03-FEB-1993) Kunio Yamane, Inst. Biological Sciences,  
Tsukuba University; Tsukuba, Ibaraki 305, Japan (Tel:0298-53-6419,  
Fax:0298-53-6006)

FEATURES  
source  
Location/Qualifiers  
1. .2739  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
292..297  
/evidence=experimental  
315..320  
/evidence=experimental  
365..697  
/gene="orf1"  
365..697  
/gene="orf1"  
/codon\_start=1  
/transl\_table=11  
/product="ORF1"  
/protein\_id="BAA2221.1"  
/db\_xref="GI:2424968"  
/translation="MSLEKTRRMNLYDFYQSLTSKQSYMPLYLDDFSLGEIAEE  
YVSRQAVDNIKRTFAMLEQIEKLLLLKFKQERKENFKLSELASGSEEEITALL  
IEALEKLD"  
711..2051  
/gene="ffh"  
711..2051  
/gene="ffh"  
/codon\_start=1  
/transl\_table=11  
/product="Ffh"  
/protein\_id="BAA21691.1"  
/db\_xref="GI:2309080"  
/translation="MAFESGLADRLQOTISKIRGKVSQDVKKMREVRLLAEADY  
NFKVVDYFKYSERAVGQVYMKSLTPGQVVKVQVEELTELMGSESKIAVAKRPPT  
VIMVGLQAGKVTTSGLANLLKKNRPKLVAAADYV RPAATKOLETIGKOLDMPV  
FSLGDVSPVETAKQAEKAKEHYDVILDTAGRLHIDHELMDELINVEIANPEI  
FLYVDSVQDAVNVKSNFNEQLGLTVLTKDGTGGAALSIRAVTPIKFAGL  
GEKLDALFHPERNAKILMGDVLTLEKQASVDDEKAKELEQMRMTSFTLDDF  
LEQIGVRNRPGLDLEQMPGAGKMKLNKTQVDEKQINHEVAIIKSMVLEKEQPD  
IINASRRKRIAGSGTSVQEVNRLKQFDEMKMKMKQTNNSKGGKFKLPTM"  
2091..2096  
-35\_signal  
2117..2122  
-10\_signal  
2157..2429  
gene  
/gene="rpsp"  
2157..2429  
CDS

/gene="rpsp"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S16"  
/protein\_id="BAA21692.1"  
/db\_xref="GI:2309081"  
/translation="NAVKIRLKRMGAKKSPFYRIVVADSRSPDRGRLETVGTYNPVA  
KPAEVLIDELALKWLQTCAPKPSDTVRNLFSSQIGIMEKFNHAKQK"  
BASE COUNT 887 a 527 c 648 g 677 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,69e-91 Length: 2739  
Score: 1647.00 Matches: 319  
Percent Similarity: 84.62% Conservative: 66  
Best local Similarity: 70.11% Mismatches: 60  
Query Match: 72.43% Indels: 10  
DB: 1 Gaps: 1

US-09-943-108A-2 (1-455) x D14356 (1-2739)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
||||| :|||||

Db 711 ATGGCATTGAGGATTAGCCGACCTGACAGCAGATTTCTAAATCCGCGAAAA 770  
||||| :|||||

QY 21 GlyLysLeuThrGluAlaAspPheLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
||||| :|||||

Db 771 GGGAAAGTCAGCGAACAGATGAAAGAGATGATGCGTGAGGTCGCTTCGCGTGCTT 830  
||||| :|||||

QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60  
||||| :|||||

Db 831 GAGGCTGACGTTAACTTTAAAGTAGTCAAGGATTTGTCAAAAAGTAAGTGAACGCGCT 890  
||||| :|||||

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLeuValGln 80  
||||| :|||||

Db 891 GTAGGCCAAGACGTCATGAAAGCTGACGCCGCCAGCAGGTTCATTAAGTTGTTCAA 950  
||||| :|||||

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
||||| :|||||

Db 951 GAGGAACGTAGCTAGCTGATGGCGGCGGAGAGACAAAATCCGCCGTCGCAAAAGGCGG 1010  
||||| :|||||

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
||||| :|||||

Db 1011 CCGACTGTTATTATGATGTCGCTCCCAAGGTGCGGTAAACGACACCAACGCGSTAAG 1070  
||||| :|||||

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
||||| :|||||

Db 1071 CTGCGCAATCTGCTGCGCAAAAAAGCATATCGCAACCGATGCTGTGTCGCGATATT 1130  
||||| :|||||

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
||||| :|||||

Db 1131 TACGCCCCAGCCGCAATTAAGCAGCTGGAAACACATCGGCAACAGCTTGATATGCTGTT 1190  
||||| :|||||

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
||||| :|||||

Db 1191 TTCTCTCTGGCGATCAGGTCACTGCTGTAGAAATAGCTAAACAGGCTATTGAGAAAGCC 1250  
||||| :|||||

QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200  
||||| :|||||

Db 1251 AAGGAAGAACATTATGACTACGTCATTTGGATACGGCGGCGCTTGCATATCGACCAT 1310  
||||| :|||||

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220  
||||| :|||||

Db 1311 GAACGTGATGATGAACATGACCAACGTCGAAGAAATCCGGAATCCGGAAGAAATTTTCCGT 1370  
||||| :|||||

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
||||| :|||||

Db 1371 GTTGTGATTCAATGACCGGTCAAGCGCTGAGATGTTGCCAAAAGCTTTAATGAACAG 1430  
||||| :|||||

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAla 260  
||||| :|||||

Db 1431 CTCGGTTTAAACCGGTGTTGTGTTGACTAGCTGGATGGAGACACACGCGCGGCGCTCGG 1490  
||||| :|||||

TITLE	The complete genome sequence of the gram-positive bacterium Bacillus subtilis
JOURNAL	Nature 390 (6657), 249-256 (1997)
MEDLINE	98044033
PUBMED	9384377
REFERENCE	2 (bases 1 to 208780)
AUTHORS	Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
TITLE	Direct Submission
JOURNAL	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchine@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48
FEATURES	Location/Qualifiers 1..208780 /organism="Bacillus subtilis" /strain="168" /db_xref="taxon:1423" 258..4588 /gene="bpr" 258..4559 /gene="bpr" /EC_number="3.4.21.-" /note="alternate gene name: bpf" /codon_start=1 /transl_table=11 /product="bacillopeptidease F" /protein_id="CAB13404.1" /db_xref="GI:2633903" /translation="MRKTKNRLISLVSTTVVISSLIPFGAAGASSVKTSPPSVKKELQ QKRASVQNSLKYTVADSSODVLKYLNTQDKDKNADQIHSYIVYNGIATVAHSAKEWKEVK VOFFPEKVLPLPKHPQLFKSPFPNNKKAQAKATGDPEENNVNDQIDAPKAWLDGLO TGVYASIDTGVEWHNPALKBEKYVGNPNPENPENENNVDPAAGEASPYDLALGG HVTPMGSEPDGTWQIGDAKWIAVAGYSEDGDTADILEAGEWLAPADKARPNP HTEWAPVYNNSGGGLSDGWTRDMVNAAARAADIFFPFSGAGNTDLFIPIGGSPITANR ANYPSFATGADINKILADFSLOGSPSYDEIKPEISAPGVNIIPSSVPQTGYDDGQV TSMAGPHVSNAALIKQNASLSVDMEILLSTAEPILDISTDPDSPNGYGHGGLXVAV FDVASVLDGRAGSLVSGSDGDDPGYQHEKVETAEYEGSDLPILTAEINDSNVTSV VLYDSIQDGTEWTAKRIISGDLHGTYQAEPIDIKGLSKYSKWIHDHDFGGHWVSSD WLDVTXPSTIQTQDETAPGQVNASCTNNNWGPVSPNTAASKGYTCGTNLTLT GNYSANAMNMPIIKAPDSGSLEFLQPKSWLNLEDPDFYGFVFLPECKNEQAGVKK YNGTSSWTDERTDISLAYKGONIQWMFNLDQSDESTAKEGWYIDDVYVLDISKAGKTVKK MKLGVEPSEKQKKPVNPKKAFSPATAVKHQNKATOPOVLPLKAQVSVVETKSTG VINKTGEPTVGASYVVEDAAEAPAMNDKGEIMLEYEGATIIVAAPGYSIDSFPS WLTGTFVKTALTKFTVGYPTGIAYDDGTAEANASYFAAGNWAATMLDAGDKGMKI TELGLRWDPDRFPDGGTEFEKVEVYATGKDGAQPKKITAGPFAEAALRGWETKVDLS SKGLMWKDDPLYVIOKSPDYPVSGLMADGTGONSRRMWOYIDKQPQGDGKADGMYMI RALVDYEAAPVELTSPTRKSTYKDNVYKGNASPGTTVHYLINGEKAEGETAAKADGT FHAGLIILKNELSTATONGTIDASSPTITVLDQKPELTDNPKDGGKTNKTCGT TVKGAADDNKLKDVAGNKKATVADGYSARILLENGRENEIKVATLDLAGNKTTKT IDVNFADPVTLSGLIEGDKNLKAGESYKTAFAESAEEDLATFTIRMPLTNARASVONAT ELPRETSPORVEGYWTATSSIIKAKGAKVEVIVRDDYDGNETRTKNGKLLMNNTEN" 4571..4598 /gene="bpt" /gene="bpr" /gene=".spoIIIGA" 4754..5683 /gene="spoIIIGA" /function="processing of pro-sigma-E (SpoIIGB) to active sigma-E (stage II sporulation)" /codon_start=1
terminator	
gene	
CDS	



QY 201 AlaleuMetAsnGluLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetIeu 220  
 DB 73753 GAACGTATGATGAATGACCAACATCGCAAGAAATCGCAATCGCAAGAAATTTCTCG 73812  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
 DB 73813 GTTGTGATTCATGACCGGTGAGACGCTGTGAATGTGGCAAAAGTTTAATGACAG 73872  
 QY 241 LeuAspValThrGlyValThrLysLysLeuAspGlyAspThrArgGlyGlyValAla 260  
 DB 73873 CTCGGTTTAACCGCTGTTGTGTGATTAAGCTGATGAGACACACGCGCGGCGTGG 73932  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerCluLysIeu 280  
 DB 73933 CTTTCTATTCCGCTGTCAACACGCAATTAAGTTTTCGAGTTTGGCGCAAAAGCTT 73992  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 DB 73993 GATCGCTAGACCGCTTCCATCTCAACGCGATGCGATCAAGGATTCGGCATGGCGAC 74052  
 QY 301 ValLeuSerIleuLeuLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspIeu 320  
 DB 74053 GTGCTGACATGATTGAAAAACACAGCCGCGCTGATGAACACAAAGCCAAAGAGCTG 74112  
 QY 321 GluLysLysMetArgGluSerPheThrLeuAspPheLeuGluLeuAspGln 340  
 DB 74113 GAACAAAAATGGAACGATGAGCTTCATGAGCATTTCTGAGCAGCTCGGCGAA 74172  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 DB 74173 GTCAGAAACATGGGCGCTGTGATGAGCTTCTGCAATGATGCGGGTSCAGGTAAATG 74232  
 QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380  
 DB 74233 AAGGCGCTGAAAAACATCCAAAGTGTGATAAAGACAGCTGAATCATGTGGAAGCAATCATC 74292  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 DB 74293 AAATCATGACTGTCTTGAAGAAGACAGCGCGATATTATCAATCCAGCGCGCGAAG 74352  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
 DB 74353 CGGATTCGAAAGAGAGCGGGACATCCGTACAGGAGCTCAACGCTGCTTAAAGCAGTT 74412  
 QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440  
 DB 74413 GATGAATGAAAAAATGATGAGCAGATGACAAACATGTCAAAAGGCAAGAAA----- 74466  
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
 DB 74467 -----AAAGGGTTTAAGTACTCTTT 74487  
 RESULT 12  
 LOCUS AP001515 303249 bp DNA linear BCT 01-FEB-2001  
 DEFINITION Bacillus halodurans genomic DNA, section 9/14.  
 ACCESSION AP001515 BA000004  
 VERSION AP001515.2 GI:12641879  
 KEYWORDS  
 SOURCE Bacillus halodurans DNA.  
 ORGANISM  
 Bacillus halodurans  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 REFERENCE 1 (sites)  
 AUTHORS Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fuji,F.,  
 Nakamura,Y. and Inoue,A.  
 TITLE An improved physical and genetic map of the genome of alkaliphilic  
 Bacillus sp. C-125  
 JOURNAL Extremophiles 3 (1), 21-28 (1999)  
 MEDLINE 99184645  
 PUBMED 10086841  
 REFERENCE 2 (sites)  
 AUTHORS Takami,H. and Horikoshi,K.

TITLE Reidentification of facultatively alkaliphilic Bacillus sp. C-125  
 JOURNAL Biosci. Biotechnol. Biochem. 63, 943-945 (1999)  
 REFERENCE 3 (sites)  
 AUTHORS Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,  
 Masui,N., Fuji,F., Takaki,Y., Inoue,A. and Horikoshi,K.  
 TITLE Sequencing of three lambda clones from the genome of alkaliphilic  
 Bacillus sp. strain C-125  
 JOURNAL Extremophiles 3 (1), 29-34 (1999)  
 MEDLINE 99184646  
 PUBMED 10086842  
 REFERENCE 4 (sites)  
 AUTHORS Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and  
 Horikoshi,K.  
 TITLE Sequence analysis of a 32-kb region including the major ribosomal  
 protein gene clusters from alkaliphilic Bacillus sp. strain C-125  
 JOURNAL Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)  
 MEDLINE 99209008  
 PUBMED 10192928  
 REFERENCE 5 (sites)  
 AUTHORS Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G.,  
 Sasaki,R., Hirama,C., Fuji,F. and Masui,N.  
 TITLE Genetic analysis of the chromosome of alkaliphilic Bacillus  
 halodurans C-125  
 JOURNAL Extremophiles 3 (3), 227-233 (1999)  
 MEDLINE 99411980  
 PUBMED 10484179  
 REFERENCE 6 (sites)  
 AUTHORS Takami,H.  
 TITLE Genome analysis of facultatively alkaliphilic Bacillus halodurans  
 C-125  
 JOURNAL Extremophiles in deep-sea environments, 249-284 (1999)  
 REFERENCE 7 (sites)  
 AUTHORS Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.  
 TITLE Replication origin region of the chromosome of alkaliphilic  
 Bacillus halodurans C-125  
 JOURNAL Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)  
 MEDLINE 99356711  
 PUBMED 10427704  
 REFERENCE 8 (sites)  
 AUTHORS Takami,H. and Horikoshi,K.  
 TITLE Analysis of the genome of an alkaliphilic Bacillus strain from an  
 industrial point of view  
 JOURNAL Extremophiles 4 (2), 99-108 (2000)  
 MEDLINE 20263314  
 PUBMED 10805564  
 REFERENCE 9 (sites)  
 AUTHORS Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T.,  
 Hirama,C., Fuji,F. and Takami,H.  
 TITLE Characterization and comparative study of the rrr operons of  
 alkaliphilic Bacillus halodurans C-125  
 JOURNAL Extremophiles 4 (4), 209-214 (2000)  
 MEDLINE 20426005  
 PUBMED 10972189  
 REFERENCE 10 (sites)  
 AUTHORS Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N.,  
 Fuji,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and  
 Horikoshi,K.  
 TITLE Complete genome sequence of the alkaliphilic bacterium Bacillus  
 halodurans and genomic sequence comparison with Bacillus subtilis  
 Nucleic Acids Res. 28 (21), 4317-4331 (2000)  
 MEDLINE 20512582  
 PUBMED 11058132  
 REFERENCE 11 (bases 1 to 303249)  
 AUTHORS Takami,H. and Takaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-NOV-2000) Hideto Takami, Japan Marine Science and  
 Technology Center, Deep-sea Microorganisms Research Group; 2-15  
 Natsushima, Yokosuka, Kanagawa 237-0061, Japan  
 (E-mail:takami@jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html,  
 URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html,  
 Tel:81-468-67-3895, Fax:81-468-66-6364)  
 COMMENT On Jan 31, 2001 this sequence version replaced gi:10174886.

```

FEATURES
    source            Location/Qualifiers
    1..303249
        /organism="Bacillus halodurans"
        /db_xref="taxon:86665"
        /note="alkaliphile"
    gene
    104..1102
        /gene="BH2267"
    CDS
    104..1102
        /gene="BH2267"
        /codon_start=1
        /transl_table=11
        /product="penicillin-binding protein"
        /protein_id="BAB05986.1"
        /db_xref="GI:10174887"
        /translation="MKNHHTIMEDWKLSTALMKKGEDIPFIASLGFANRABRIPNE
        HHTFGLASCKLFTALICQLVERAGKLSFDTPLSDWLDPFPNTVTHLLTHTSGVP
        DYDEETDDFDLNDKDMYHLRLKDFLPFOHAPMKFPFGPHFYNNAGTILGL
        VVESGVTFQEVYANFYVQFAGHESGYFAFTLPKATLGYIDLEGGWKNLYSL
        PVISSGGAYVTAEDMKLWALMRHELLNETYTKLLTPHVCDDDDYGVGVWIK
        QODGATSKVMGYDPGVCFSAFVPTSGIVVVCANQSSGAYDVMAAIALISEA"
        complement(1146..1469)
        /gene="BH2268"
        complement(1146..1469)
        /gene="BH2268"
        /note="BH2268"
        unknown"
        /codon_start=1
        /transl_table=11
        /protein_id="BAB05987.1"
        /db_xref="GI:10174888"
        /translation="MKVITLQGSTKFRKQFEQASVLTILKGHVLSVGFQESGIEI
        TEEQOVFENIHRKTIENSGDEIFVLDVGYIGRSTRKEIFATSKGMPRIYSESLA
        STFS"
        complement(1636..2373)
        /gene="BH2269"
        complement(1636..2373)
        /gene="BH2269"
        /note="BH2269"
        unknown"
        /codon_start=1
        /transl_table=11
        /protein_id="BAB05988.1"
        /db_xref="GI:10174889"
        /translation="MVKKGLTVGVFALSTGLMACNGQBEVEKNIDGASGEIPSI
        BEIDINLPDPVATNGESIDKETVYVLEQQQWMAOYGISLGEKGOMLSLEE
        QTIFLNLQLOQATKEIEVSEIDEELAFYVAQSSSEALMRALERGGSDDE
        LREIEHYVQKIYVEETEINATYEEIEIQARYEEKEQYITTEDELPFTELQSNIEQR
        LIAKEQEQLVLFKEKLRDSDITVHI"
        complement(2582..3367)
        /gene="BH2270"
        complement(2582..3367)
        /gene="BH2270"
        /codon_start=1
        /transl_table=11
        /product="ABC transporter (ATP-binding protein)"
        /protein_id="BAB05989.1"
        /db_xref="GI:10174890"
        /translation="WADILQVEQVTLRKAKKPILOQISNOVEQGEHVALVINGSGKT
        SULLIHTGWPTEGHVAVLGNRGYHPIQEVKRGVWYSMSLDRFHTKQDTVLEI
        VLSGLHTGVEPTGEADVAKARALETGLTHLTDRLPFLTSQGERKRAFLARAW
        AKPELFDIPEPTGLDILAROLLQTLLESVATTKGAPTEFLYTHYPEIQPFTIHLV
        LKDGHVTAQGRKDEVLTKEMITAFAFGCSLDVIAQSGRYWVKPN"
        complement(3378..4310)
        /gene="BH2271"
        complement(3378..4310)
        /gene="BH2271"
        /codon_start=1
        /transl_table=11
        /product="dipeptidase"
        /protein_id="BAB05990.1"
        /db_xref="GI:10174891"
        /translation="MKFIDHCDALLKLLWEDPAREYTDPAIEANAVLMRMEGVRAQC
        FAIFVGNIPFDKFOVALQAELEFHEKILKHPQSFIERWDDLLLTLDADQIGAILT
        LEGADPFQDDPTKLTLYRLVRSRLGTWNQANLCADGVGPGRGGLTETWKGAVVSLN
        NEHVLTDVSHLSERGFWDVLEYAOFPAVTHSNAKAICKHRENTLDQLKALIEAGGF
        IGIVFPHLFTGTDDASIDDVKKHIEHICSLGCKHVGSGDFDGIQCFVDGLEHAGR
        FQWFMDELLKRYDEELVRGFANDNFKRATPISV"
        4406..4840
            /gene="BH2272"
        CDS
        4406..4840
            /gene="BH2272"
            /codon_start=1
            /transl_table=11
            /product="heat shock protein class I (low molecular
            weight)"
            /protein_id="BAB05991.1"
            /db_xref="GI:10174892"
            /translation="MSREESKSLPNVYQDFLRSIDDFSKSQHLDHNGMFAPSIP
            IRYEKGDFLIEALPGVSKNQIQLDVFPQYVIRIKHEEVLREMRDEKGIYSKIS
            SQAREVIFPPVQEGDIKARLNGLLTITAPNHRKRINIE"
            4864..5193
                /gene="BH2273"
            CDS
            4864..5193
                /gene="BH2273"
                /note="BH2273"
                unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB05992.1"
                /db_xref="GI:10174893"
                /translation="MINLHPQWTWSQAGYQQPQHSWPNHPQPTSSGGQMYTGTG
                PSFPFPFPPYFQSKPKASPKAAFTNGDFTDINKTFTQVDQVYVTVNQVPLVKS
                VGSIFIK"
                complement(5202..5510)
                /gene="BH2274"
                complement(5202..5510)
                /gene="BH2274"
                /note="BH2274"
                unknown"

Alignment Scores:
Pred. No.:      2,43e-85      Length:      303249
Score:          1584.00      Matches:      307
Percent Similarity: 83.33%      Conservative: 73
Best Local Similarity: 67.32%      Mismatches: 70
Query Match:     69.66%      Indels:      6
DB:              1          Gaps:      2

US-09-943-108a-2 (1-455) x AP001515 (1-303249)
Qy 1 MetAlapheGluglyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||
Dy 210434 ATGGCGTTTGAAGAGTTAGCGGCGGCTGCAAGATACGTTGAAGAAATTCGCGGAAA 210375
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||
Dy 210374 GGAAGAGTAAGTGACAGACGCTAAAGAAATGATGCGTGAAGTTCGTGCGCTCCCTA 210315
Qy 41 GluAlaAspValAsnAspPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
|||||
Dy 210314 GAAGCGGATGAACCTTAAAGTGTGAGCAATTCATCGCCATGTCGAGGAAAGCC 210255
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||
Dy 210254 CTCGGTCAAGAAGTCATGAAAGCTGACACCTGACACAGGTCATCAAGTCGTCAT 210195
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||
Dy 210194 GAAGAGCTGACGGCGCTGATGGAGGCGAGCAAGAAATTCGGTTCGAGCAAGCCG 210135
Qy 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||
Dy 210134 CCACCGGTCGTGATGATGCTCGGTTCACGGGGCTGGGAAACACGACGACGCGGAAA 210075
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||

```

Db 210074 CTTGCCAACCATTTGGGAAAAAGACAAATCGGAACCGGATTCTCGTGGCGCGATATT 210015  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 210014 TATCGTCGGCTCGGATTAAACACCTTGAGACATTAGTAAACAATTAACATCGCTGTG 209955  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 209954 TTTAGCCTGGTGATCAAGTAGCCCTGTGTAATTGCGAAACAGCGCATCGACGGCG 209895  
 QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGln 200  
 Db 209894 AAAGACAGCATCACCATATGTGCTCATGACAGCTGGACGCTTCACATCGATGAA 209835  
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 209834 GAGCTGATGATGAATCGCAACAAGTAAGAAAGTGGCAACCGCATGAAATCTTGCTC 209775  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlnSerPheAspAspGln 240  
 Db 209774 GTCTTGATCGGATGACAGGGCAGATGCGGTTAATGTGCGGAAAGCTTCATGAGCAG 209715  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 209714 CTCGATGTTTCGGCGCTGTTTTCAGGAAGCTGGATGGGATACTCGAGGGGAGCTGCC 209655  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 Db 209654 CTCFCGTTAAGCGGTTACGAACACGCCGATCAAAATTTGCCGGATGGGTGAAGAAGATT 209595  
 QY 281 AspGlyLeuGlnLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 209594 GACCAGCTCGAACCTTCCACCAGAGCGAATGGCTTCGGGATCTTGGGATGGGGAT 209535  
 QY 301 ValLeuSerLeuIleGlnLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
 Db 209534 GTGCTGACATCTATGAAAAGACAAATCCACTGCGATGAAGAAAGAAAGCGCGTGAACCT 209475  
 QY 321 GlnLysLysMetArgGlnSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340  
 Db 209474 GAGAAAAAATCGGTACAATGACATTACGTTTCGATGATTCTTGGAGCAACTCGACCAA 209415  
 QY 341 ValLysAsnLeuGlyProLeuAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 209414 GTGCGGAGCATGGGACCATCTCATGACTGCTCGGATGATGCCAGGATGAACAAGCA 209355  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGlnLysGlnIleAspHisIleLysAlaIleIle 380  
 Db 209354 AAAGCCTTGAAAACTTGCAAGTGCAGCAGAGAACAATTAGCCGGCTTGAGGCGATTGTC 209295  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 209294 GTTCCATGAGCAAGCAAGAAACAACATCTAGCATCAATTAATGCGAGTCGTCGCA 209235  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
 Db 209234 CGGATCGGAAAGAGAGCGGGCGCGGATTCAAGATGTCACCGGCTGTAAACAGATT 209175  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGly---GlyGlyLysGlyLysGly 439  
 Db 209174 GAAGAGATGAAAAGATGATGAAGCAATGACGAACATGACGAGCAAGCAAGCAAGAA 209115  
 QY 440 LysArgAsnGlnMetGlnAsnMetLeuLysLysMetAsnLeuProPhe 455  
 Db 209114 GGGAAAGGC-----TTTGGAGGCTTTAAATGCGCTTT 209082  
 RESULT 13  
 AX433432 AX433432 1136 bp DNA linear PAT 28-JUN-2002  
 LOCUS  
 DEFINITION Sequence 1847 from Patent WO0291113.  
 ACCESSION AX433432  
 VERSION AX433432.1 GI:21658236  
 KEYWORDS

SOURCE Bacillus licheniformis.  
 ORGANISM Bacillus licheniformis  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 REFERENCE 1  
 AUTHORS Berka,R. and Clausen,I.G.  
 TITLE Methods for monitoring multiple gene expression  
 JOURNAL Patent: WO 029113-A 1847 11-APR-2002;  
 NOVOTRIZ Biotech, INC. (US) ; NOVOTRIZ A/S (DK)  
 FEATURES  
 source location/Qualifiers  
 I.1136  
 /organism="Bacillus licheniformis"  
 /db\_xref="taxon:1402"  
 BASE COUNT 338 a 263 c 326 g 209 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4,68e-77 Length: 1136  
 Score: 1402.00 Matches: 270  
 Percent Similarity: 87.30% Conservative: 60  
 Best Local Similarity: 71.43% Mismatches: 48  
 Query Match: 61.65% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-943-108A-2 (1-455) x AX433432 (1-1136)  
 QY 1 MetAlaPheGluGlyLeuSerGluArgGluAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGCATTCGAAGGATTAATGCCGACCTGACGACCAACAGATCTCAAAATCCGCGGAAA 60  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaPhe 40  
 Db 61 GGAAGGTAAACGAGCAAGCTGCAAGAGATGATGCGGAAATCCGTTAGCGCTCCT 120  
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 121 GAAGCGAGCTCAATTTAAATCTGTAAGATTTGTGAAAAGGTAAAGCAACGAGCT 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 Db 181 GTCGGACAGACGATTATGAAAGCCCTGACGCCGACGAGCATTAATAAGTGGTAAA 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 241 GAAGAGCTGACGAGCTGATGGCGGCGAGGAAAGATTTCCGCTGCAAAAGCGTCG 300  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CGGACCGCTCATCATGATGTCGGTCTGCAAGGGCGCGTAAACAGACGTCGACCGGAAG 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysProMetLeuValAlaAlaAspIle 140  
 Db 361 CTGCGCAACCTGTTCCGAAAAACAATAACCGCAATCCGCTGCTGCGCAGCAGCATC 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 421 TACCGCGCGGCGGCGATCAACAGCTGGAAACGCTGCGCAAGCAGCTCATATGATGCGGTA 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTTTCACTGGGTGACAAAGTCTAGTCCGTTGGAGATTGCCAAGCAGGCCCATCGAAGGCG 540  
 QY 181 LysGlnGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGln 200  
 Db 541 AAAGAGATGATCAACGATACCTGATCTCATGATGACGGGCGGACCGCTTCATATGATGAA 600  
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 GAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlnSerPheAspAspGln 240  
 Db 661 GTCGTTGACTCCATGACGGGTCAAGACGCGGTCAATGTTGCCAAAAGCTTTAAACGACAG 720

Qy 241 LeuaspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaala 260  
Db 721 CTCGGTGTGACAGGTCTCTGACGAGCTTGACGCGTACTAGAGCGGAGCGGCC 780  
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280  
Db 781 TTGTCAATCCGCGCGGTGACAAACACGCGCATCAAAATCGCGGAATGGCGAAAGCTT 840  
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
Db 841 GATGCGCTGTAGCGCGTTCATCTCGAACGAGTGGCTTCGCGGATACGCGAATGGCGCAT 900  
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
Db 901 GTGCTCACCTTATCGAGAAAGCCAGCCACAGTAGACGAGAGAAAGGCGAAGGACTC 960  
Qy 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
Db 961 GAGCAAAAATGAAAAATATGAGCTTACGCTCGACGACTTTTGGAAACAGCTCGGACAG 1020  
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
Db 1021 GTCCGCAATATGGCGCGCTTGAGACCTGATCAATGATGATGATGATGATGATGATG 1080  
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAla 378  
Db 1081 AAAGGCTTAAAAAGCTGAAGTTGAGTAAACACAGCTCACCATATCGAAGCG 1134  
RESULT 14  
AE007684  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TATSOV,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,  
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.  
Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum  
J. Bacteriol. 183 (16), 4823-4838 (2001)  
21359325  
11466286  
REFERENCE  
2 (bases 1 to 11024)  
Childress,D., Zeng,Q. and Smith,D.R.  
Direct Submission  
Submitted (24-JUL-2001) GTC Sequencing Center Production,  
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100  
Beaver Street, Waltham, MA 02453-8443, USA  
FEATURES  
source  
location/Qualifiers  
1. 11024  
/organism="Clostridium acetobutylicum"  
/strain="ATCC 824"  
/db\_xref="ATCC:824"  
/db\_xref="taxon:11488"  
84. 317  
/product="Acyl carrier protein, ACP"  
/gene="acpA"  
84. 317  
/gene="CAC1747"  
/codon\_start=1  
/transl\_table=11  
/product="Acyl carrier protein, ACP"  
/protein\_id="AAK79713.1"  
/db\_xref="GI:15024716"  
/translation="WVFEKVKDIIADQLGIDATEIKNESSFIDDLGADSLDIYELIMA

LEEFEDIEMPEAEKVKSSVGVVYNIKAHTEE"  
444. 1136  
/gene="CAC1748"  
/note="rncs"  
444. 1136  
/gene="CAC1748"  
/codon\_start=1  
/transl\_table=11  
/product="dsRNA-specific ribonuclease III"  
/protein\_id="AAK79714.1"  
/db\_xref="GI:15024717"  
/translation="MEERLEELLESKLGTEFONINLLVTALTSSVANENKNAEYNE  
RLFLGDAVLQLSISEYFFKYPPTISGGLTKRALVTCGMSLSIGERQLCKYIRM  
SHEELTGGTRVYSIIADCEAVAAIYIDKGFDAKFIREFEGTQIQAVERNKILL  
DAKRLQELQSGKGTDIKTLVRHEGPPHRRKFFVNLNFDNDVKSTGEGITTKDAEQ  
DAAKALKGLDN"  
1137. 2213  
/gene="CAC1749"  
1137. 2213  
/gene="CAC1749"  
/codon\_start=1  
/transl\_table=11  
/product="BioB/LipA-like protein"  
/protein\_id="AAK79715.1"  
/db\_xref="GI:15024718"  
/translation="MGKSHYIPIFVPEHCGHDCVFCNONTITGESKYNVDYVKST  
VEDYLDIPNKSATIEISFGGTFATINIEKOKELISVABVKIEIGIKYIRLSTPDD  
YIDEILTNLKNYSVDIIELGVQSLDSEVLLASGRGHTKEDYVAKSLIKKYGTGLGH  
QIMGLPGDNAEKDIETTRVICMKPNICRIYPALVIKDTAWERKMYRNGIYKPYLEE  
CVISIKVFGLFANDIEVIGLQPTTEINTDELVAGPFPAPRFLVESGSMNEMI  
FNLKCESTVDIYNPKDISKLYSNKKTFFYSMKDKLNLNLRINVINQNEVEKDTLK  
MNFDDCKIVSKYEVLVKYKEG"  
2270. 2530  
/gene="CAC1750"  
2270. 2530  
/gene="CAC1750"  
/codon\_start=1  
/transl\_table=11  
/product="SpoVS-related protein"  
/protein\_id="AAK79716.1"  
/db\_xref="GI:15024719"  
/translation="MEVLKVSQPKSVAGALAAVLNRSNSAEVQAVGAGAINQAVK  
AIAITGTFVAPNGIDLVVVPAPAEISTEGEDRTAIFKFIYESR"  
2747. 6322  
/gene="CAC1751"  
/note="smc"  
2747. 6322  
/gene="CAC1751"  
/codon\_start=1  
/transl\_table=11  
/product="Chromosome segregation SMC protein, AtPase"  
/protein\_id="AAK79717.1"  
/db\_xref="GI:15024720"  
/translation="MFLASIRGKFSADKTDLIFKNGITAVVPGNSGKSNISDAV  
LWVGQSVKNLGRGMEDVIFAGTQYKRSVGLAQVSLIDNSDKQNLNDYSEVTVSR  
KLVSQSYIYNNTCKRLKDIDELFMDTGIGEGYSIIQGGKIEAVLSGKIPERRAL  
LEEAAGIVKFKRYDAEKLENTQNLVINDILFTEERLEPRLSESKAKRPFEL  
SDELKTEINTIYISDINDYKNDLQKRMADIKSIDENYKDEKISLAKSTVATSEL  
DEFDAKVSNNKTYVESKSEHOKLSEIELLEKTSNDAKNKLYKEIDOLNSIVN  
LKSRVETOLKTDKYNKELLSKINKSEKKNDIGLJREWEKSKOYKNDADIDII  
STSQNNNEVILKKEISNESKESLEKIRAGEGYSKSLKINEVTKNLSIELVKINDK  
ISGYNQIRENSKISKLNRILISDEKLNELNSKNLEANKMLINLEKQYEGYNR  
SVKNMQHVTGFDVGPSSFVLGEIVKVKFEFETAVEISLGAISDITLDDNTAK  
KLINLKSNGLRATFLPLNIQKRLINISDATRHEKFGITASELIDYDSTPLAVN  
VYLGRTVIVDNDMSALKATKLSYSPKIVTLGTEVNVPGSLTGTSTYSKAAISIGRK  
RETEELNLNNVSALEQSNKILLENKKVVELNCLDLTDIHEKTELKIKER  
LKSIDISEKLNKSYNTSVGIGTFIKINKHLEKLVKEENKALKRANNNLID  
ELERKLKDNSKVLNNEEIMSKVDKAKSDENLMSSTRETERVKVHMNMENKILSL  
KNEIDFENQVILANKTYENNVQIQVIELENSFKDSEVERSKAGTENTKRN  
SLOGINVLQKLETKHYEINLAKIETESNLYRLNSDFKLITYSAAAEKFEIEM  
ISVKKRTIDELKREINKMGVNVASIEYKVEKVTWNSQKEDLNKDELIRVLEE  
WTKSKRTVTFNFKKLNFENKIFRDLFGKGNADLILSGDDELSSLEINVEPPGKL  
QNINLMSGGEGKLSAIALLLFALKKPKTPFCFILDIEAALDDANVARYAEFLRFSEN

```
gene
6352..7263
/feature="CAC1752"
/feature="fts"
6352..7263
/feature="CAC1752"
/codon_start=1
/transl_table=11
/product="signal recognition particle GTPase, FtsY"
/protein_id="AAK79718.1"
/db_xref="GI:15024721"
/translation="MLGNFEDKLKNGLSKTNKLTSTKISDWLSTVITIDDLVLEEE
ILITSDIGVETSLYIIDKLKEIKENKQKPSLVNDCLKEIKIDLGDKGSGMPKQI
PTILVGVNGVGTTSIGKISSKSNLKVIMAAADTPRAAIDOLEWSKRANVD
LIKHEGSDPAVVDIAEAKKADVLICDTAGKLNKLNMLEIKNLIIDREF
SOSTRTLLVDATGQNAVIAQFNEVCKVGVGLVILKIDGTANGGIVIFIKHOLDI
PVEFVGVGEGIDDOEFNSDFVEALF"
7341..7691
/feature="CAC1753"
7341..7691
/feature="CAC1753"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein, YLXM
B subtilis ortholog"
/protein_id="AAK79719.1"
/db_xref="GI:15024722"
/translation="MEBRIYLSMLSTYGSILTEKQVNMRLYDDDLSPETAEINN
TRQAHLDIKRKMLNLYENKLVKRYKNEDIKDKIDKILYILKIDKQDKNVQ
LIDEIKDINSFSI"
7710..9059
/feature="CAC1754"
/feature="ifh"
7710..9059
/feature="CAC1754"
/codon_start=1
/transl_table=11
/product="signal recognition particle GTPase Pfh"
/protein_id="AAK79720.1"
/db_xref="GI:15024723"
/translation="MAFEGSLTKLOAMKLRGKLSKDIKDAMEVKLALLRADV
NYIKNFVYVGVKCLGNEVMSLTPGOQVIKYNEELTNMGKESEKTEFENGIT
VIMVGLGAGTTCMGKLSLKKKPLVACDIYRPAAIKOLEVCKSDYVPVF
SMGDKNPVDSKAMKAKENGILNVIIDTAGRLHIDDQLMNELENKSEFNKEIL
LVDSMTGDDAVNVAESFDKKBELTGVLTLDGTGCGAALSIREMTGKPKYVGLG
ERNDDLEIFHPDRMASRLGMDVLTLEKQASIDERQAKELGDRMLSOEFDDFL
QAFQMKLIGPTGLLEKLVFNSMLKGVDSKNEGEMKYEAIKSMYAKERNPS
LITSTARRKRRIALGSGTTVQEVNKLKLNFEQMKMKQPKGNKFSKGLFGKMPF"
9097..9342
/feature="CAC1755"
/feature="rpsp"
9097..9342
/feature="CAC1755"
/codon_start=1
/transl_table=11
/product="Ribosomal protein S16"
/protein_id="AAK79721.1"
/db_xref="GI:15024724"
/translation="MKVIRKMGAKKAPFYRIVVADSRSPRDKFTTEELGYNPTTE
PVTFKVDADKVNAMKNGAOPSETVKLLDKSGVTTK"
9363..9590
/feature="CAC1756"
9363..9590
/feature="CAC1756"
/codon_start=1
/transl_table=11
/product="Predicted nucleic acid binding protein,
containing KH domain"
/protein_id="AAK79722.1"
/db_xref="GI:15024725"
/translation="MKQLLETIAKSLVDCPDEVQVSEVTGQSIIELKVPEDMGKV
IGKQRIAKAIRTVIKAAVAKENKRVVVEII"
9718..10218
gene
```

```

/feature="CAC1757"
9718..10218
/feature="CAC1757"
/codon_start=1
/transl_table=11
/product="RimM protein, required for 16S rRNA processing"
/protein_id="AAK79723.1"
/db_xref="GI:15024726"
/translation="MEDFFSIGQILINHGVRGELKIYPLTDDINRFDLDSVYVONEI
KVVISVKQPKNLKLEIGDITLDEAVKYNKIKVLRDVAELKESQIFIKDIIICGN
VFEDNKDLGEYDVIYVSTKNNDYVIRKEGQDILVPALKDILVKIEINKKIVIKAV

Alignment Scores:
Pred. No.: 2..75 Length: 11024
Score: 1393.00 Matches: 280
Percent Similarity: 76.47% Conservative: 71
Best Local Similarity: 61.00% Mismatches: 94
Query Match: 61.26% Indels: 14
DB: 1 Gaps: 5

US-09-943-108a-2 (1-455) x AE007684 (1-11024)

QY 1 MetalPheGluGlyLeuSerGluGluGluGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 7710 ATGGCTTTTCGAGGGATTATCTACAAATTAACAGCAGCAATGAAAACTCAGAGGAAAA 7769

QY 21 GlyLysLeuThrGluAlaAspLysLysLysLysLysLysLysLysLysLysLysLys 40
Db 7770 GGTAAACCTTCGAAAAAGCATATAAAGATGCTATGAGAGAGATTAAGCTTCTCTTTA 7829

QY 41 GluAlaAspValAsnPhelLysValLysGluPhelLysThrValSerGluArgAla 60
Db 7830 GAAGCAGATGTTAACTACAAATAGTTAAGACCTTCTTAAAGTAGTAGGAGAAAGTGT 7889

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysLys 80
Db 7890 CTAGTAATGAAGTATGGAAGCTTACTCTCGGCAGCAAGCTTATAAAAAATAGTTAAT 7949

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysAsnMetSerAsnLys 100
Db 7950 GAAGAGTTTAACCTACCTCACTGAGGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8009

QY 101 ProThrValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 8010 ATTACTGTTAATGAGGAGTGGACCTCAAGCTCGAGGAGGAGGAGGAGGAGGAGGAGG 8069

QY 121 LeuAlaLeuLeuMetArgLysLysLysLysLysLysLysLysLysLysLysLysLys 140
Db 8070 CTTTCTCTTACTTTAAAAAGAAAAA--AATAAAAAAGCCACTTCTTGTGCTGATGATATA 8136

QY 141 TyrArgProAlaAlaAlaAsnGlnLeuGlnThrValGlyLysGlnLysLysLysLys 160
Db 8127 TATAGCAGCAGCAGCTATAAGCAGGTAGGTAGGTGGAAGGATCAATAGATGCTCCTGT 8186

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnValThrAsnAlaLeuLysHisAla 180
Db 8187 TTTTCTATGGGGGATAGGTTAATCCTGTTGATATTCAAGGCTGCAATGAGCATGCT 8246

QY 181 LysGluGluHisLeuAspPheValLysLeuAspThrAlaGlyArgLeuHisLysLys 200
Db 8247 AAGAAGAAATGGTTTAAATGTAGTTATATAATAGACTGCGGAGAACTTTCATATTGATG 8306

QY 201 AlaLeuMetAsnGluLeuLysGluValLysLysLysLysLysLysLysLysLysLys 220
Db 8307 CAGCTTATGAATGAACCTGAGAACATATAAAGTGAAGTAACTAAGGAAATTTTGCTT 8366

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 8367 GTTGTGATTTATGACAGGTCAGATGCGTAATGTTGCTGAACCTTTGATATATAG 8426

QY 241 LeuAspValThrGlyValThrLeuThrLysLysLysLysLysLysLysLysLysLys 260
Db 8427 CTTGAACCTTACTGCTGTTGCTTACTAACTGTGATGATGATGATGATGATGATGATGAT 8486
```







[illegible]

Search completed: February 25, 2003, 02:59:00  
Job time : 4087 secs

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

PD 17-MAR-1999.  
 XX PF 24-AUG-1998; 98EP-0306741.  
 XX PR 10-SEP-1997; 97US-0927216.  
 XX PA (SMIK ) SMITHKLINE BEECHAM.  
 XX PI Wallis NG;  
 XX WPI; 1999-169238/15.  
 DR P-PSDB; AAY00910.  
 XX  
 New Staphylococcus aureus Signal Recognition Particle (SRP) with  
 PT protein (fih) and RNA (fifs) components - the SRP gene and protein  
 PT useful as diagnostic reagents and for prevention and treatment of  
 PT Staphylococcal infections which cause otitis media, septic arthritis  
 PT and toxic shock syndrome  
 XX  
 XX Claim 1: Page 28; 35pp; English.  
 XX  
 CC This sequence encodes the Staphylococcus aureus signal recognition  
 CC particle (SRP) fih component. Fih polynucleotides are useful for  
 CC diagnosing a disease related to expression of fih polypeptides by  
 CC analysing for the presence/amount of fih protein in a sample due to  
 CC infection of a micro-organism with the gene, or determining the nucleic  
 CC acid sequence encoding fih. Fih polypeptides and antagonists are useful  
 CC for treatment of an individual in need (polypeptide) of, or needing to  
 CC inhibit (antagonist) fih polypeptide levels. Fih polypeptides and  
 CC polynucleotides are useful for identifying agonists and antagonists by  
 CC binding and observing the affect of fih polypeptide activity, which are  
 CC potential anti-microbial agents. Fih polypeptides and antigenic fragments  
 CC are also useful for inducing an immunological response (T cell/antibody)  
 CC to protect against disease, by direct administration (vaccine), or via a  
 CC vector (gene therapy). Anti-fih antibodies are useful as antagonists, and  
 CC for protecting against disease. Diseases diagnosed, prevented and treated  
 CC include those caused by infection, especially bacterial infection,  
 CC including otitis media, conjunctivitis, toxic shock syndrome, wound  
 CC infection and septic arthritis. Fih polypeptides are antimicrobial, and  
 CC are useful for bathing wounds and implants prior to surgical  
 CC implantation.  
 XX  
 SQ Sequence 1368 Bp; 521 A; 186 C; 288 G; 373 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 3,47e-197 Length: 1368  
 Score: 2274.00 Matches: 455  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 20 Gaps: 0  
 US-09-943-108A-2 (1-455) x AAX27221 (1-1368)  
 QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCATTTGAAGCGTTATCAGAACCCCTGCAGCGACGATGCAAAATAATGCGTGAAG 60  
 QY 21 GlyLysLeuThrGluAlaAspPheLysLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 61 GGTAACCTTACTGAAAGCTGATATTAAGATAATGATGCGTGAAGTAAGATTAGCGTTATT 120  
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheLysThrValSerGluArgAla 60  
 Db 121 GAGCGTGCAGTAACCTTAAGTGGTAAAGAAATTTATTAACAGCTATCAGACGCCGA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLeValGln 80  
 Db 181 TTAGGTTCCGATGTAATGCAATCATTAACACCGGCGCAACAAGTTATTAATAAGTTCAA 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 241 GAGGAATTACGAAGTTGATGGGTGGAGAAATATCATCGATTATATATGTCATAATAACCA 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCTACTGGTTGGTTATGATGGTTGGTTTACAGGTGCTGGTAAACACAACTCAGAGTAAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
 Db 361 TTAGCATTATTGATGCGTAAANAATACACAAAAACCTATGTTAGTTGCGAGAGATATT 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGGAACAATATGATATTCCTGTA 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TAGAGTGAAGGAGATCAAGTAAGGCCACACAAATGTAACTAATGATTCATTAACCATGCT 540  
 QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
 Db 541 AAGAGAGACATTTAGACTTTGTAATCATTGATACAGCAGTCGATTACACATCGATGA 600  
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 GCATTGATGACGAATTAAGAAGTAAAGAAATGCTAAACCAACGAAATATATGTTA 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
 Db 661 GTTGTCGATCAATGACGGTCAAGATGCTCAATGTTCAGAAATCTTTTGACGATCAA 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 721 CTTGATGTCACAGGTGTTTACCTAACTAAATTAGATGCTATACACGTGCTGTCAGCT 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 Db 781 TTATCTATTCTCGGTGACACAAAAACCAATTAATTTGTTGATGATGAGTAAAGTTA 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 GATGGTTAGAGCTATTCCATCTCCTGACGATGAGCATCACGTATTTTAGGTATGGGTGAT 900  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 Db 901 GTGTTAAGTTTAATGAAAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960  
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
 Db 961 GAGAAAAAGATGCGTGAGTCATCGTTTACTTTTAGATGATTTTTTAGAACCACTTGATCAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 1021 GTGAAAAATCTAGGACCACTGGATGATATATGAAATGATTCAGGATGATGAATAAATG 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
 Db 1081 AAAGGCGTAGATAAGCTTAATATGAGTGAAGCAAAATGATGATATTAAGCGATATC 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 CAGTCAATGACCGCGGTGAAAGAAACAATCCAGACACATTGGAATGTATCAGCTAAAAAG 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 CGTATTGCTAAAGGCTCGTGGCTTCATTACAGAGGTCATCGTTTGATGAACAATTT 1260  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440  
 Db 1261 AACGATATGAAGAAAAATGATCAACAAATTCACCTGCTGCGGTGAAGGTAAGGTAAGGTAAG 1320  
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
 Db 1321 CGCAATCAATGCAAAATATGTTAAAGGATGATGAATTTACCGTTT 1365

## RESULT 2.

NTS011 2  
AAS55331

ID AAS55331 standard: DNA: 1368 bp.

[illegible]

AC AAS55331;

XX

DT 13-FEB-2002 (first entry)

XX

DE Staphylococcus aureus

XX

KW Antisense; d

KW antibiotic; antibacterial; d

Category	Sub-category	Value
C	C1	1
	C2	1
D	D1	1
	D2	1
E	E1	1
	E2	1
F	F1	1
	F2	1
G	G1	1
	G2	1
H	H1	1
	H2	1
I	I1	1
	I2	1
J	J1	1
	J2	1
K	K1	1
	K2	1
L	L1	1
	L2	1
M	M1	1
	M2	1
N	N1	1
	N2	1
O	O1	1
	O2	1
P	P1	1
	P2	1
Q	Q1	1
	Q2	1
R	R1	1
	R2	1
S	S1	1
	S2	1
T	T1	1
	T2	1
U	U1	1
	U2	1
V	V1	1
	V2	1
W	W1	1
	W2	1
X	X1	1
	X2	1
Y	Y1	1
	Y2	1
Z	Z1	1
	Z2	1

Staphylococcus aureus.

XX  
DN  
WC000170055-22

PN  
YY  
W0Z001/0953-AZ.

AA  
PD  
27-SEP-2007

ED 21-SEP-2001.  
XX

21-MAR-2001: 2001WO-US09180-

[illegible]

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-

PR 26-MAY-2000; 2000US-

PR 23-OCT-2000; 2000

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.  
PA  
XX  
XX  
XX  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
FI Yamamoto RT, Xu HH;  
XX  
XX  
DR WPI; 2001-6111495/70.  
DR P-PSEB; AAU37472.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
PT  
XX  
XX Claim 27; Seq ID No 8968; 51lpp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence encodes an essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part of the published sequence listing.

CC of the print.

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 1368 BP; 517 A; 187 C; 291 G; 373 T; 0 other

Alignment Scores:

Pred. No.:	5.26e-196	Length
------------	-----------	--------

Qy	1	MetAlaPheGlnGluLeuSerGluuA-gLeuGlnAlaThrMetGlnLysMetArgGlyLys	20
Db	1	ATPGCAATTCAAGGGTTATCAGAACGCTTGCAGCGAGATGCAAAAATACGCTGGTAAAG	60
Qy	21	GlyLysLeuThrClnAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
Db	61	GGTAACCTTACTCAAGCTGATAAAGATAAAGATGAGTGAAGATTAGCGTTACTT	120
Qy	41	GluAlaaspValAsnPhelYsValValLysGluPheIleLysThrValSerGluArgAla	60
Db	121	GAGCTGACCTAAACTTTAAGTGTAAAGAAATTATTAAACACAGTATCAGACGCGCA	180
Qy	61	LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln	80
Db	181	TTAGGTCTCCAGTGTAATGCAATCATTAACACAGCGGCAACAAGTTATTAAATAGTCTCA	240
Qy	81	AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
Db	241	GATGAATTACCGCATGTTGATGGGTGAGAAAATACGTCGATTAATATGTCAAATPAACCA	300
Qy	101	ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys	120
Db	301	CTTACTGTTCTGATGATGGTTTACAGGTGCTGGTAAACACACACACACACACACAC	360
Qy	121	LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle	140
Db	361	TTAGCATTTATTGATCGTAAAAAATACAAAAAACCTATGTTAGTTCGACGAGATATT	420
Qy	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
Db	421	TATGTCACGACGCGATAAATCAATTACAAACAGTAGGGAACAATATGATTTCTCTGTA	480
Qy	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
Db	481	TACAGTGAAGAGATCAAGTAAAGCCACACACAAATGTTACTTAATGCAATTAACACAT	540
Qy	181	LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu	200
Db	541	AAAGAAGAACAATTAGACTTTTGTAATCATTTGATACAGCAGGTGATACACATCGATGAA	600
Qy	201	AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu	220
Db	601	GCATTTGATGAACCAATTAAAGAAAGTAAAGACATTCGTAAACAAACGAATATTGTTA	660
Qy	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln	240
Db	661	GTTGTCAATCAATGACGGTCAAGATGCTGCATGTTGCAGATCTTTTGACGATCAAC	720
Qy	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla	260
Db	721	CTTGATGTCACAGGTGTTACCTTAACATAAATTAGATGGTGATACAGGTGGTGCAGCT	780
Qy	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
Db	781	TTATCTATTCTGCTGGTGACACAAACCAATTAATTTGTTGGTATGAGTGAAGAAATGA	840
Qy	281	AspGlyLeuGluLeuPheHisProGluuA-gMetAlaSerArgIleLeuGlyMetGlyAsp	300
Db	841	GATGTTTAGAGCTATTCCTTCCTGAACGTATGGCATCAGTATTTAGGTATGGGTGAT	900
Qy	301	ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGlnLysAlaLysAspLeu	320
Db	901	GTGTTAGTTTATTGAAAAGCCACACAGATGTGGATCAAGAAAAGCAAAAGATTTA	960
Qy	321	GluLysLysMetArgGlnSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln	340
Db	961	GAGAAAAAGATGCGTGAGTCATCTCTTACTTTTAGATGATTTTTTAGAACCACTTGATCAG	1020
Qy	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
Db	1021	GTGAAAAATCTAGACCACTGGATGATATTATGAATATGATTCAGGATGAATAAATG	1080

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
Db 1081 ARAAGGCTAGATAAGCTTAATATGAGTGAAGAAAGCAAAATGATCATATTAAGCGATATTC 1140  
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
Db 1141 CAGTCAATGACCGCGCTGAAGAAAGCAATCCAGACACATTTGAATGTATCACGTAAAAAG 1200  
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
Db 1201 CGTATTGCTAAGGGTCTGGTCGTCATTACAAGAGTCATTCGTTGATGAACAAATTT 1260  
QY 421 AsnAspMetLysGlyMetMetLysGlnPheThrGlyGlyLysGlyLys 440  
Db 1261 AACGATATGAAGAAATGATGAACAGTTCACCTGGTGGCGGTAAAGGTAAAAAGGTAAA 1320  
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProphe 455  
Db 1321 CGCAATCAAAATGCAAAATATCTTAAAGGTATGAATTTACCGTTT 1365  
RESULT 3  
AAS51783  
ID AAS51783 standard; DNA; 1365 BP.  
XX  
AC AAS51783;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Staphylococcus aureus DNA for cellular proliferation protein #200.  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Staphylococcus aureus.  
XX  
PN W0200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
DR WPI; 2001-611495/70.  
DR P-PSDB; AAU33924.  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
PS Claim 27; Seq ID No 4365; 511pp; English.  
XX

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC  
CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 1365 BP; 518 A; 190 C; 285 G; 372 T; 0 other;

Alignment Scores:  
Pred. No.: 1,2e-193 Length: 1365  
Score: 2235,00 Matches: 447  
Percent Similarity: 98,90% Conservative: 3  
Best Local Similarity: 98,24% Mismatches: 5  
Query Match: 98,28% Indels: 0  
DB: 23 Gaps: 0

US-09-943-108A-2 (1-455) x AAS51783 (1-1365)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
Db 1 ATGGCATTTTCAAGGCTTTATCAGAACGCCCTCAAGCGACGATGCAAAAATTCGGTGGTAA 60  
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
Db 61 GGTAACTTACTGAAGCTGATATAAGATAATGTCGTGAAGTAAGTAGCGTTACTT 120  
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
Db 121 GAGGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAAACAGTATCAGAACGCGCA 180  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
Db 181 TTAGGTTCCGATGTAATGCAATCATTAACACCAGGCAACAAGTTATTAAATAGTTCAA 240  
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
Db 241 GATGAATTAACGAGCTGATGGGTGGAGAAATATATCGATTATATGTCATTAACCA 300  
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
Db 301 CCTACTGTTGTTATGATGGTTTACAAGGTGCTGGTAAACCAACAACACTGCAGGTAAA 360  
QY 121 LeuAlaLeuLeuMetArgLysLysValAspLysProMetLeuValAlaAlaAspIle 140  
Db 361 TTAGCTTTATTTGATCGTAAAGAAATACACAAAACCTTTAGTTGCAGCAGATATT 420  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
Db 421 TATCGTCCAGCAGCGATTAATCAATTACAACAGTAGGGAACAATTTGATATTCCTGTA 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAATAATTTGTAATTCATTAACATGCGATGAA 540  
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
Db 541 AAAGACACACATTAGACTTTGTAATCATTTGATACAGCAGGTCCGATTACATCGATGAA 600  
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
Db 601 GCATTGATGAATGAATTAAGAAGATTAAGACATTCGTAAACCAACCAAGAAATATTGTTA 660  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
Db 661 GTTGTGCTGATTCATGACGGGTCAAGATGCTGCAATGTTGCAGAAATCTTTTGACGATCA 720  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260  
Db 721 CTTGATGTCACAGGTTTACCTTAACATAAATAGATGGTGATACACGTTGGTGGCGAGCT 780

```

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValClyMetSerGluLysLeu 280
DB 781 TTAATCTATTCGTTGCGTGACAAACCAATTAATTTGTTGGTATGAGTGAACATTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
DB 841 GATGGTTTAGAGCTATTCCTCCTGAACGATGCGATCATCGTATTTTCAGGCGCTGGGTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
DB 901 GTGTCAAGGTTAATTGAAAGACCGACAGATGTGATCAAGAAAGACCAAGATTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
DB 961 GAGAAAAAGATGCGCGAGTCATCATTTACTTTAGATGATTTTTTAGAACAACTTGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
DB 1021 GTGAAAAAATTAGGACCACTGGATGATATTTATTAATAATGATTCAGGTATGAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
DB 1081 AAAGCTTAGATAGCTTAAATGAGTGAAGAAAGCAATGATCATATTAAGCCATATTC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
DB 1141 CAGTCAATGACGCGCGCTGAAAGAAACAATCCAGACACATTAATGATTCACGTAAAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
DB 1201 CGTATTGCTAAAGGTTCTGTCGTTTCATTAACAAGAGTCAATCGTTTGTGAACAATTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys 440
DB 1261 AACGATATGAGAAATGATGAAACATTCACCTGCGCGGTAAAGGTAAAGGTTAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
DB 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365

RESULT 4
AAS54393
ID AAS54393 standard; DNA; 1368 BP.
XX AC AAS54393;
XX DT 13-FEB-2002 (first entry)
XX DE Staphylococcus aureus DNA for cellular proliferation protein #705.
XX KW Antisense; ds; prokaryotic cellular proliferation gene;
XX KW antibiotic; antibacterial; drug design.
XX OS Staphylococcus aureus.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;

```

```

XX WPI; 2001-611495/70.
DR P-PSDB; AAU36534.
XX
PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
PS Claim 27; Seq ID No 8030; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1368 BP; 520 A; 190 C; 285 G; 373 T; 0 other;

Alignment Scores:
Pred. No.: 1,21e-193 Length: 1368
Score: 2235.00 Matches: 447
Percent Similarity: 98.90% Conservative: 3
Best Local Similarity: 98.24% Mismatches: 5
Query Match: 98.28% Indels: 0
DB: 23 Gaps: 0

US-09-943-108A-2 (1-455) x AAS54393 (1-1368)
QY 1 MetalAPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB 1 ATGGCATTGAAGCGCTTATCAGAACGCGCTGCAAGCAGCATGCAAAAAATGCGTGGTAAG 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
DB 61 GGTAACTTACTGAAGCTGATATAAGATATGATGCGTGAAGTAAAGATTACGGTTACTT 120
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
DB 121 GAGCGTGACGTAACTTAAGTGTAAAGAAATTTATTAACACGATATCAACAGCGCCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
DB 181 TTAGGTTCCGATGTAATCAATTAACACAGGCAACAACTATTATTAATAATAGTTCAA 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
DB 241 GATGAATTAACGAAGTTGATGGGTGGAGAAATATATCGATTATATGCAATAAACA 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
DB 301 CTAATCTGTTGTTATGATGCTGTTTACAGGTGCTGGTAAACACAACTCCAGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrsLysLysProMetLeuValAlaAlaAspIle 140
DB 361 TTAGCATTTATTGATGCGTAAAAAATACAAACAAAACCTATGTTAGTTGCACGACATAT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
DB 421 TATCTCCAGCAGCGATAAATCAATTAACAACAGTAGGAAACAAATTTGATTCTCTGTA 480

```



QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisIala 180  
 Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAACAATGTPAATCAATCAATCAATCAATGCT 540  
 QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
 Db 541 AAAGAAGAACATTTAGACTTTGTAATCAATGATACAGAGTGTGATACACATCATGATGAA 600  
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 GCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
 Db 661 GTTGTCGATTCATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaIa 260  
 Db 721 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 Db 781 TTATCTATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 GATGTTTAGAGCTATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
 Db 901 GTGTCAGGTTAATGAAAGCCGCAACAGATGTCGATCAGAAAGCAAGCAAGCAAGCA 960  
 QY 321 GlnLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
 Db 961 GAGAAAAGATGCGCGAGTCATCATCTTACTTTAGATGATTTTGTAGACACTTGTATCAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 1021 GTGAAAATTTAGGACCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
 Db 1081 AAGGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 CAGTCAATGACGCGCGCTGAAAGCAACATCCAGACACATTCATGATGATGATGATGAT 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 CGTATTCGTAAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440  
 Db 1261 AAGCATATGAGAAATGATGAAACATTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
 Db 1321 CGCAATCAATGCAAAATATGTTAAAGGATGATGATGATGATGATGATGATGATGATGAT 1380  
 RESULT 5  
 ABQ70511/c  
 ID ABQ70511 standard; DNA; 2256 BP.  
 XX  
 AC ABQ70511;  
 XX  
 DT 29-AUG-2002 (first entry)  
 XX  
 DE Listeria monocytogenes 4b contig DNA sequence #453.  
 XX  
 KW Antibacterial; Listeria; food contamination; mutational analysis;  
 KW infection; ds.  
 XX

OS Listeria monocytogenes 4b.  
 XX  
 PN WC200228891-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 04-OCT-2001; 2001WO-FR03061.  
 XX  
 PR 04-OCT-2000; 2000FR-0012697.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Kunst F, Glaser P;  
 XX  
 DR WPI; 2002-332479/37.  
 XX  
 PT New genomic sequences from Listeria species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators  
 XX  
 PS Claim 14; SEQ ID 3324; 180pp; French.  
 XX  
 CC The present invention relates to nucleic acid sequences  
 CC (ABQ7188-ABQ71212) from Listeria sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of Listeria (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of Listeria (potential therapeutic agents), also for  
 CC treating infections by Listeria, and are useful as immunogens in  
 CC anti-Listeria vaccines.  
 CC  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2256 BP; 533 A; 471 C; 420 G; 832 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1,45e-146 Length: 2256  
 Score: 1719.50 Matches: 331  
 Percent Similarity: 86.81% Conservative: 64  
 Best Local Similarity: 72.75% Mismatches: 55  
 Query Match: 75.62% Indels: 5  
 DB: 24 Gaps: 1  
 US-09-943-108A-2 (1-455) x ABQ70511 (1-2256)  
 QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1666 ATGCGCATTTGAAGGACTAGCTGGAGAGACTCCCAAGAAACAATGAACAATAATCGCGCAAA 1607  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 1606 GGAAGAGTAAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1547  
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 1546 GAAGCGGATTTAATCTTTAAAGTCGTTAAACAATTTATTAACAACAGTAAGCAAGCTGCT 1487  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 Db 1486 GTCGCGCGGAGCTTATGAAAGCCTTAACACCGCGCAACAAGTTATCAAAATTCGTTCAA 1427  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 1426 GAAGAATCAACAGCTTAATGGCGGAGAGAAAGAAATTCGGAACAGCGGACGCGCA 1367  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 1366 CCACGCGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1307  
 QY 121 LeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140



```
Db 188164 GAAGAATTAAACAGCCTCATGGCGGCGAAGAAAGTAAAAATCGGAACAGCTGATCGCCG 188223
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 188224 CCACCGTTATATGATGGTAGGTGTACAGAGCTGGTAAACAACCACTTCAGGAAA 188283
Qy 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysPrometLeuValAlaAlaAspIle 140
Db 188284 CTGCCTAATTATTACGTAAATAATATATCGTAAACCTTTACTAGTCGACGAGATAT 188343
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 188344 TACCGACCTGCAGCAATCAACAATAGAAACACTGGCAACAATTAGATATGCCAGTA 188403
Qy 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180
Db 188404 TTTTCTTAGGGGATTCAGTAGCCCAAGTAGAAATCGGGAACAGCTATCGCTAAAGCA 188463
Qy 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 188464 AAAGAAGAACAATTAGATTATGTCAATTCGATACAGCTGGCTCTTCATATCGACGAA 188523
Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 188524 ACTCTGATGGAGCAATTAACAAGTAGGAAGAAATCGCTACGCCCACTGAAATTTACTT 188583
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 188584 GTAGTTGATCAATGACTGGCAGAGCGCAGTAATGTGGCCCAAGCTTCAACGACAA 188643
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 188644 TTAGAANTTACCGGCGTGTATTAAACAATAAGTAGACGTGATACACGTGGTGGGCGACA 188703
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 188704 CTTTCCATCCGTTCACTAGTCAGCAAGAAACCAATCAATTCGTTGCTACCGGTGAAAAATG 188763
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 188764 GAAGCAATCGAAACCTTCATCCGATCGTATGGCTTCAGAAATTCGGCATGGGTGAT 188823
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 188824 GTACTTTCCTTATTGAAAAAGCACAACTGATGTAGATGCAGAAAAAATGAAGACTATG 188883
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
Db 188884 GAACAAAAAATGAAGACACAGCATGACCTTAGATGACTTCTTGGACCAATTCGAACAA 188943
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 188944 GTAAACAATGGGACCACTAGTAACACTATAAATAATGATGCCAGGGGCAACACAAATG 189003
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 189004 AAAGGCTCGACACATGCAAGTAGATATAACAACCTCGGTACATCGTCAAGGCGATAAT 189063
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 189064 AAATCCATGACCAAAAACAGAAAGATATAATCCGGACATCATCAATGCGACGACAGAAA 189123
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 189124 CGAATTCCTCGTGGAGCGGACGCCCAATTCAGAAATCAATCGCTCCCTTAACCAATTT 189183
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
Db 189184 GCTGAAATGAAAAAATGATGACGAAATGACTGGTGGGGAAGGTAAGAAGGTAA 189243
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 189244 -----AATCCATTCGGCAATTTCAAAATGCCATTT 189273
```

```
RESULT 7
ABN92495
ID ABN92495 standard; DNA; 1014 BP.
XX AC ABN92495;
XX DT 24-JUL-2002 (first entry)
XX DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1958.
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy; gene; ds.
XX OS Staphylococcus epidermidis.
XX PN US6380370-B1.
XX PD 30-APR-2002.
XX PF 13-AUG-1998; 98US-0134001.
XX PR 14-AUG-1997; 97US-055779P.
XX PR 08-NOV-1997; 97US-064964P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX DR WPI; 2002-381255/41.
XX DR P-PSDB; ABP39950.
XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX PT polypeptide, useful for diagnosing and treating bacterial infections -
XX PS Disclosure; SEQ ID 1958; 267pp; English.
XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
XX CC antibacterial activity and can be used in gene therapy. The sequences
XX CC can also be used in the diagnosis and treatment of bacterial infections,
XX CC particularly S. epidermidis infections. The sequences can be used to
XX CC screen for compounds able to interfere with the S. epidermidis life
XX CC cycle or inhibit S. epidermidis infection.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC USPTO web site.
XX SQ Sequence 1014 BP; 354 A; 156 C; 221 G; 283 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 3,75e-130 Length: 1014
Score: 1534.00 Matches: 307
Percent Similarity: 97.29% Conservative: 16
Best Local Similarity: 92.47% Mismatches: 9
Query Match: 67.46% Indels: 0
DB: 24 Gaps: 0
```

US-09-943-108A-2 (1-455) x ABN92495 (1-1014)

```
Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 16 ATGCGATTTGAGAGGATATCCGATCCGTTACAGCACGATGCAAAAATCGCGTATAA 75
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 76 GGAAGAGTAACAGACAGATATTAACACGATGATGCGTGAAGTGAGATTAGCGTTATG 135
Qy 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 136 GAAGCGGATGTTAACTTCAAGATTGTTAAGGAATGTTTAAGAATGTTTCAGAACCGAG 195
```



```
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 241 GAAGAATTAAAGAAACGTTAGTTCAGAAACGGTTGAACATAATCTCCAAATC 300
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 101 ProThrValValMetSerValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 301 CCACAGTGTATGATGACAGGGTTACAGGGCTGTGAACAACTTTTACTGTAAA 360
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 361 TTACGAAACACCTTAATGAAACTGAACCGTCGTCGCTTTTATCGCTGTCGCT 420
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 421 TATCGTCCAGCAGCATTGATCAGTTGAAGTTTGTAGTCAACAATTAGAAGTCCCGTT 480
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 481 TTTGATATGGGACAGATGCTTAATCCAGTGGAAATGTCTCAAGGTTTAGCATAGCA 540
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 181 LysGlnGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 541 AAGAAAGAAAGAAATGATGCTTAAATGATACGGCCGCTTTACATTTAGCAAG 600
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 601 GCTTTAATGACGAATTCAGAAATTAAGAGTTGGCTAATCCCAATGAATTCGTGA 660
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlnSerPheAspAspGln 240
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 661 GTTGTGTGATCGCATCGCGGACAGATGCTGCTCAACGTTGCAGATAGTTTAAATGAACAG 720
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyGlyValAlaAla 260
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 721 CTGGATTAACGCGGTTGTTATTCACAAATGACGCGGTACTCTGGGGGGGCTGCG 780
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 781 CTGTCAATTCGGGACAGTAAACGGCGCTCCGATTAATTTGCTGTTGTTGAAATTA 840
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 841 ACCGATTTAGAAATTTCCATCCGATCGTATGCGATGCTATCTCTAGTATGGGGAC 900
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 301 ValLeuSerLeuIleGluAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 901 ATGTGACGCTAATTTGAAAGGCGCAACAGATTACGATGAGAAAAAGCAGAGCACT 960
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 961 GCTCAAAAATGAAAGAAACAGTTTTCATTTAAGCATTTTCATTGACCAATTGGATCAA 1020
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 341 ValLysAsnLeuGlyProLeuAspIleMetLysMetIleProGlyMetAsnLysMet 360
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 1021 GTATGGCATGGCCAGTGAAGACTTATTAANAATGATCCCTGGAATGATACAG 1080
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 361 LysGlyLeuAspLysLysAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 1081 CCTGGTATTGAAATGTCAAAGTCCATCCAAAGATGTGGCAGAAACGGCGGATGTC 1140
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 1141 CTATCAATGACCCCGCAGACGTAATAATCTGATTTAATAATCTGATCAATCAATTT 1260
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 1201 AGAATTCAGCTGGTTCAGGAATAGTCTGTTGAAGTCATCTGATGATTAACAACTTT 1260
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 421 AsnAspMetLysLysMetMetLysGlnPheThr 431
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
Db 1261 AAAGAATCCAAAAAATGATCAACAAATGTCCAAAGGGGATGATCAACATTCCTGGTATG 1320
   :::::::::::::: ::::: ||| ::::: ||| ::::: |||
QY 432 -----GlyGlyGlyLysLysLysGlyLysArgAsnGlnMetGlnAsn 446
```

```
Db 1321 GATCAATCTAGTGGCGGTTAAGCAAGTTAGTAAATG---GCCATGAATCT 1377
   ||||| ||||| ||||| |||||
QY 447 MetLeuLys 449
   ||::: |||
Db 1378 ATGATGAAG 1386
   ||::: |||
RESULT 9
AAS53127
ID AAS53127 standard; DNA; 1434 BP.
XX
AC AAS53127;
XX
DT 13-FEB-2002 (first entry)
XX
DE Enterococcus faecalis DNA for cellular proliferation protein #555.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Enterococcus faecalis.
XX
PN W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PE 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-287931P.
PR 16-FEB-2001; 2001US-269308P.
XX
(PA ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
WPI: 2001-611495/70.
P-PSDB; AAU35268.
XX
New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids -
Claim 27; Seq ID No 6764; 51lpp; English.
XX
The invention relates to antisense inhibitors of genes essential to
prokaryotic cellular proliferation, their use in identifying the
genes, their use in the discovery of novel antibiotics, the essential
genes themselves and the encoded proteins. The prokaryotes used are
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
invention is also useful for the identification of potential new targets
for antibiotic development. The antisense nucleic acids can also be used
to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1434 BP; 500 A; 230 C; 340 G; 364 T; 0 other;
XX
Alignment Scores: 1.77e-119 Length: 1434
Pred. No.:
```

Score: 1418.50 Matches: 273  
 Percent Similarity: 77.54% Conservative: 86  
 Best Local Similarity: 58.96% Mismatches: 89  
 Query Match: 62.36% Indels: 15  
 DB: 23 Gaps: 2

US-09-943-108A-2 (1-455) x AAS53127 (1-1434)

QY 1 MetAlaPheGluGlyLeuSerGluArgGluGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCTTTTGAGAGTTTAAACAAACCGCTTACACAGCAAGTAAATCCGTCGTAAAG 60  
 QY 21 GlyLysLeuThrGluAlaAspPheLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 61 GGAAGAAGTTCCGAGCGCGAGCTAAAGAAATGATGCGAGAAATCCGTTGCTTTATTA 120  
 QY 41 GluAlaAspValAsnPheLysValLysValLysGluPheLeuLysThrValSerGluArgAla 60  
 Db 121 GAAGCCGAGCTTAATTTACAAGTGGTCAAGATTTCAAAAACGCGTCAGAGAAAGGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLeuValGln 80  
 Db 181 GTAGAGTCGAGTATTAGAAACCTTATCACGCCCCACAAATGTAAATTTGTTGAT 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysLeuMetSerAsnLysPro 100  
 Db 241 GAAGAAATTAACGAAACGTTAGGTTTCAGAAACGGTTGAACCTGAATAAATCTCCAAAATC 300  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlnAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCGACAGTGATATTGATGACAGGGTTACAAAGGGCGTGTAACAACACTTTTACTGGTAAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
 Db 361 TTAGCAAAACACTTAATGAAACCTGAACACGCTCTCGCTTTTAATCCGCTGCTGACGT 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnLysLeuAspIleProVal 160  
 Db 421 TATCGTCCAGCAGCGATTGATGATGAGTTTGAAGTTTGTAGTCAACAATTAGAAGTTCCCGTT 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLysValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTTGATATGGGAACAGATGCTAATCCAGTGGAAATTTGTTCTCAAGGGTTAGCATTAACA 540  
 QY 181 LysGluGluHisLysAspPheValIleLeuAspThrAlaGlyArgLeuHisLysLeuGlu 200  
 Db 541 AAAGAAAGAAATGATGATGCTTAATGATACGCCGCCGCTTTACACATTTAGACAA 600  
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLeuLysProAsnGluLysLeuMetLeu 220  
 Db 601 GCTTTAATGGCAATTAACAAATAATTAAGAGTTGGCTAATCCCAATGAATTCGTGTA 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
 Db 661 GTGTGTGATGCGATGACGGCGCAAGATGCTGTCACACGTTCAGATGTTTATGAACAG 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 721 CTGGAATTAATGGGGTTGTTATTACAAATGGACCGGCTACTCGTGGGGGGCTCG 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280  
 Db 781 CTGTCAATTCGGCAGTAACGGCGCTCCGATTAAATTTGTCGGTTCGTGTGAAAAATTA 840  
 QY 281 AspGlyLeuGluPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 ACCGATTTAGAAATTTCCATCCCGATCGTATGTCGAGTCGTATCTAGTATGGGGAC 900  
 QY 301 ValLeuSerLeuLysLeuGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 Db 901 ATGTGTAGCGCTAATTTGAAAAAGCGCAACAGATTACGATGAGAAAAAGCAGCAAGAACTT 960  
 QY 321 GluLysLysMetArgGluSerPheThrLeuAspPhePheLeuGlnLeuAspGln 340

Db 961 GCTCAAAAATGAAGAAACACGTTTGTGACTTTAACGATTTTCATTCGCAATTCGATCAA 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspPheLysMetLysMetLysMetLysMetLysMetLysMet 360  
 Db 1021 GTTATGGCAGTGGGACGATTAAGACTTATTAAAAATGATCCTCGAATGAGTAACATG 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnLysPheHisLysLeuLysAlaIleIle 380  
 Db 1081 CTGTGATTTGAATATGTCAAAGTCGATCCAAAGATGTCGACGCAACGGCGGATGGTC 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 CTATCAATGACCCCTGCAGAACCTGAAATCTCTGATCTATTAAATCTAGTCGCGTGGC 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnLysValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 AGAATTCAGCTGGTTCCAGGAATGTTGTGTTGAAGTCATCTGATGATTAAACAATTT 1260  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThr----- 431  
 Db 1261 AAGATATCCAAAAAATGATGCAACAAATGTCCAAGGGGATATGACATTCCTGGTATG 1320  
 QY 432 -----GlyGlyGlyLysGlyLysGlyLysGlyLysArgAsnGlnMetGlnAsn 446  
 Db 1321 GATCAATGCTAGTGGCGCGCTTAAAGCAAGTTAGTAAATG---GCCATGAATCGT 1377  
 QY 447 MetLeuLys 449  
 Db 1378 ATGATGAAG 1386  
 RESULT 10  
 AAXI2974  
 ID AAXI2974 standard; DNA; 6729 BP.  
 XX AAXI2974;  
 XX 19-MAR-1999 (first entry)  
 XX Enterococcus faecalis genome contig SEQ ID NO:37.  
 XX Enterococcus faecalis; contig; detection; Enterococcal infection;  
 KW vaccine; attenuation; computer readable medium; ds.  
 XX Enterococcus faecalis.  
 XX WO9850555-A2.  
 XX 12-NOV-1998.  
 XX 04-MAY-1998; 98WO-US08985.  
 XX 14-NOV-1997; 97US-0066009.  
 PR 06-MAY-1997; 97US-0044031.  
 PR 16-MAY-1997; 97US-0046655.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Barash SC, Dillon PJ, Kunsch CA;  
 WPI; 1999-045171/04.  
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 PS Claim 1; Page 406-409; 2084pp; English.  
 XX A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC AAXI2938 to AAXI3919 represent these nucleotide sequences which are  
 CC primary nucleotide sequences, also known as contigs. The computer-based

Db	4442	CTTGGAATTACTGGGGTTGTTATTACCAAAATGGATGGCGATCTCTGCTGGGGGGGCTGGC	4501
Qy	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
Db	4502	CTGTCAATTCGGGCGAGTAAACGGGTGCTCGGATTTAAATTTCTGGTCTGGTGAAAAATA	4561
Qy	281	AspGlyLeuLLeuPheHisProGluArgMetLsSerGileLeuGlyMetGlyasp	300
Db	4562	ACCGATTGAGAAATTTTCCATCCCGATCGTATGTGCGAGTCGTACTAGTGATGGGGAC	4621

[illegible]

DD 40ZZ AIVTIVACGCIHAIIGAAAAAGGCCACACAGGATACGATGAGAAATATAGCAAGAAATACII 4001

[illegible]

DD 474Z GITAUGGCAUGGGACCGATTGAAGACITAIATAAAATGATCCCTGGAAATGAGTAAACATG 480U

QY 361 LysGlyLeuAspLysGluAsnMetSerGluLysGlnIleAspHisIleuLysAlaIlelle 380  
|||:||||: : : : ||| : : : |||:||||: : : |||:||||: :  
DB 4802 CCTGGTATTGAATAATCTCAAGTCGATCCAAAGATGTGCACGGAACCGGGCATGGTC 4851

[illegible]

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGluValAsnArgLeuMetLysGlnPhe 420

DD	4922	AGAAATTCAGCTGGTTTCAAGGAAAATAGTGTTGTGAAGTCAATCGTATGATTAAACAATTT
QV	421	AaAaScnMetIvsIrsMetMetIvsClnpbemPhy-----

Db 4982 AAGCAATCCAAAAAATGATGCAACAATGTCCAAAGGGGATATGAACATTCCTGGTATG 5041

[illegible]

Qy 447 MetLeuLys 449

Db 5099 ATGATGAAG 5107

ABK74556  
ID ABK74556 standard; DNA; 1136 bp.

AC ABK74556;  
XX

XX DE Bacillus licheniformis genomic sequence tag (GST) #1847.

Differential gene expression; genomic sequenced tag; GST;  
altered culture condition; environmental stress;  
KW KW

XX	Physiologische provocation, us.
XX	
OS	Bacillus licheniformis.

XX  
PN  
XX  
WC200229113-A2.

PD 11-APR-2002.  
XX  
XX  
BE 05-OCT-2001: 2001WQ-UG21437

XX 06-OCT-2000; 2000US-0680598.  
PR 07-MAR-2001; 2001US-0705250.  
EE 07-MAR-2001; 2001US-0705250.

XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
2

XX Berka R, Clausen IG;  
 PI WPI; 2002-416684/44.  
 XX Monitoring differential expression of several genes in first Bacillus  
 PT cell relative to expression of same genes in one or more second  
 PT Bacillus cells, by using substrate containing Bacillus genomic  
 PT sequenced tag array -  
 XX Claim 4; SEQ ID NO 1847; 200pp; English.  
 XX The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions.  
 CC environmental stress or other physiological provocation. Extensive  
 CC follow-up characterisation is unnecessary, when one spot on an array  
 CC equals one gene or one open reading frame, since sequence information is  
 CC available. This sequence represents a genomic sequence tag (GST) used in  
 CC the method of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1136 BP; 338 A; 263 C; 326 G; 209 T; 0 other;

Alignment Scores:  
 Pred. No.: 4.18e-118 Length: 1136  
 Score: 1402.00 Matches: 270  
 Percent Similarity: 87.30% Conservative: 60  
 Best Local Similarity: 71.43% Mismatches: 48  
 Query Match: 61.65% Indels: 0  
 DB: 24 Gaps: 0

US-09-943-108A-2 (1-455) x ABK74556 (1-1136)

QY 1 MetalaphedGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCATTCGAGGATTAGCCACCGACTGCGCAACAGATCTCAAAAAATCCCGGAAAA 60  
 QY 21 GlyLysLeuThrGluAlaAspLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 61 GGAAGAGTAACGGAGCAAGACGCTCAAGAGATGATGCGGAAGTCGCTTACGCTCCTT 120  
 QY 41 GluAlaAspValAsnPhelYsValValLysGluPhelLysThrValSerGluArgAla 60  
 Db 121 GAAGCGGACGTCATATTTAAATCTCGTGAAGATTTTGTGAAAAAGGTAAAGCGAGAGCT 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlnGlnValLysLysLeuValGln 80  
 Db 181 GTCGGACAGAGCTATGAAGACCTGACCGCCGACAGACAGATCAATTAAGTGTAATA 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysAsnMetSerAsnLysPro 100  
 Db 241 GAAGAGCTGACGAGCTGATGGCGGCGGAGAAAGCAAGATTTCGCCCTCAAGAGGTCG 300  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CGGACCGTCATCATGATGTCGTCTGCAAGGGGCGCGTAAACACGACGTCGACCGGAAG 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspLe 140

Db 361 CTGTCACCACTGCTTCGAAAAAATACCGCAATCCGCTGCTGCGCACGACATC 420  
 QY 141 TyrArgProAlaAlaLysGlnLeuGlnThrValGlyLysGlnLeuAspLeuProVal 160  
 Db 421 TACCGCGCGGCGGATCAACACAGCTGGAAACGCTCGGCAAGCAGCTCGATATATCGGTA 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTTTCTACTGGGTGACAGGTGCTCGGTGGAGATTGCCAAGCAGCGCCATCGAAAGCG 540  
 QY 181 LysGluGlnHisLeuAspPheValLysLeuLeuAspThrAlaGlyValGlnLeuHisLeu 200  
 Db 541 AAAGAGATCATACAGCTACAGCTTCATCATGATGCGGCGGACCGCTTCATATTGATGAA 600  
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeu 220  
 Db 601 GAGCTGATGGATGAGCTGAGCAGGTAAAGAAACCGCTCAGCCGGAAGAGATTTCCTT 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
 Db 661 GTCGTGACTCCATGACGGGTCAAGACGCGGTCAATGTGCCAAAAGCTTTAACGAACAG 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260  
 Db 721 CTCGCTGTCACAGGTGTCATCTTGACGAGCTTGACGCGGATCTAGAGCGGACGCGCC 780  
 QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280  
 Db 781 TTGCAATCCGCGCGCTGACAAACACGCGCATCAAAATCGCGGAATGGCGGAAAAGCTT 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgLeuGlyMetGlyAsp 300  
 Db 841 GATGCGCTGGAGCGCTTTCATCTGACGCGATGCTGCGGATCTCGGAATGGCGGAT 900  
 QY 301 ValLeuSerLeuLeuGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
 Db 901 GTGCTCACCTTATCGAGAAAGCCCGACGCCAAGCTAGACGAAAGAGCGGAGGACTC 960  
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340  
 Db 961 GAGCAAAAAATGAAAAATATGAGCTTTACGCTCGACGACTTTTGGAAACAGCTCGGACAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspLeuMetLysMetLeuProGlyMetAsnLysMet 360  
 Db 1021 GTCCCAATATGGGCGCGCTTGAGACGCTGATTCAAATGATGCCCGCGCAGCGCAAAATG 1080  
 QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnLysPheHisLeuLysAla 378  
 Db 1081 AAAGCTTAAAAAAGCTGAAGGTTGAGTAAACACAGCTCAGCCATATCGAAGCG 1134  
 RESULT 12  
 AAH90837  
 ID AAH90837 standard; DNA; 1605 BP.  
 XX  
 AC AAH90837;  
 XX  
 XX 02-OCT-2001 (first entry)  
 XX  
 DE 2CFE 28 coding sequence.  
 XX  
 KW Antibacterial; vaccine; gene therapy; bacterial cell wall viability;  
 KW CFE; CEG; Conserved essential gene; bacterial infection;  
 KW antisense therapy; antibiotic resistance; ds.  
 XX Streptococcus pneumoniae.  
 XX WC200149721-A2.  
 XX  
 PD 12-JUL-2001.  
 XX  
 PF 29-DEC-2000; 2000WO-US35604.  
 XX



PR 30-DEC-1999; 99US-0174089.  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Dougherty TJ, Pucci M, Dougherty BA, Davison DB, Bruccoleri RE;  
 PI Thanassi JA;  
 XX WPI; 2001-496721/54.  
 XX  
 XX Nucleic acids encoding conserved essential genes involved in bacterial  
 PT replication which are potential targets for the treatment of antibiotic  
 PT resistant bacterial infections -  
 XX  
 PS Claim 30; Fig 49; 380pp; English.  
 XX  
 CC The present invention relates to nucleic acids (AAH90701-AAH90918)  
 CC encoding polypeptides (AAH01002-AAH0114), which are essential for the  
 CC viability of a bacterial cell wall. The acronym CFE stands for "CEG for  
 CC Expression", where CEG detects for "Conserved Essential Gene". The nucleic  
 CC acids are useful for detecting the presence of proteins essential for the  
 CC viability of a bacterial cell wall in samples such as cells, tissues,  
 CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,  
 CC and for detecting corresponding target nucleic acid molecules with  
 CC complementary sequences. The nucleic acids are also useful for  
 CC determining whether a genomic nucleotide sequence of interest is  
 CC essential for viability of a bacterial cell or whether it resides within  
 CC an operon, by integrating an exogenous nucleotide sequence comprising a  
 CC portion of an open reading frame of the genomic sequence of interest  
 CC (comprising 200-500 base pairs) into the genomic sequence of interest  
 CC which confers a selectable phenotype to the cell, and determining cell  
 CC viability with a selection agent such as chloramphenicol. The nucleic  
 CC acids and proteins are also useful as vaccines and for treating bacterial  
 CC infections with gene therapy and antisense therapy. The nucleic acids  
 CC also enable identification of targets suitable for the treatment of  
 CC antibiotic resistant bacterial infections.  
 XX  
 SQ Sequence 1605 BP; 484 A; 317 C; 393 G; 411 T; 0 other;

Alignment Scores:  
 Pred. No.: 9.54e-110 Length: 1605  
 Score: 1312.00 Matches: 254  
 Percent Similarity: 74.95% Conservative: 90  
 Best Local Similarity: 55.34% Mismatches: 101  
 Query Match: 57.70% Indels: 14  
 DB: 22 Gaps: 2

US-09-943-108a-2 (1-455) x AAH50837 (1-1605)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCATTTGAAAGTTTAAACAAACGTTTCGAGACGTCCTTTAAAAATCTACGTAAAAA 60  
 Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 61 GGAATAATCTGATCTGATGTCGACAGGCAACCAAGAAATTCGCTGGCCTTCCTC 120  
 Qy 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 121 GAGCGCGAGCTTGCCCTGCTGTTGTAAGGACTTATCAAGAAGTTAGTGAGCGTGA 180  
 Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 Db 181 GTGGCGCATGAGGTGATGATCACTTAATCTGCGCAACAGATATTAAAAATCGTTGAT 240  
 Qy 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 241 GAGGAAGTACAGCGCTTTAGTTCGTACGCGCAGAAATATCAAGTCACTTACATGAT 300  
 Qy 101 ProThrValValMetMetValGlyLeuGlnGlyLysThrThrThrAlaGlyLys 120  
 Db 301 CCAACCATCATGATGTTGGTTTACAAGGGCTGGTAAACACACCTTTGCTGCTA 360  
 Qy 121 LeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140

Db 361 TTGCCCAACAACACTCAGAAGAGAAATGCTGCTCTTGTATGGTGGCGGGATATT 420  
 Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 421 TATCGTCCAGCTGCCATTGACCACTTAAGACCTTGGGACAAACAGATTGATGCTGCTGC 480  
 Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTGGCACTTGGACAGAGAACAGTGTGAGATTGTACCTCAGGTTTGGAGCAAGCC 540  
 Qy 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGln 200  
 Db 541 CAAACTAATCATACAGCATATGCTGTGATTGATACTGCGGCTCTTTCAGATTGATGAG 600  
 Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 CTCCTCATGAATGAGCTTCGTGATGTGAAGCATTTGGCTCAACCAATGAATGCTGCTT 660  
 Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
 Db 661 GTCGTTGATGCTATGATTGCTCAGGAGAGCAACCAATGTTGCGCTGAGTTTAAATGCTCAG 720  
 Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 721 TTGGAAGTGACTGGGTCATCCTTACCAAGATTGATGCGATACCTCGTGTGTGCTGCT 780  
 Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 Db 781 CTCCTGTTGCTGCATCTACTGGAACCAATCAAGTTTCACCTGTCAGGTGAAAGATT 840  
 Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 ACGGACATTGAAACCTTCCACCAGACCGCATGCTACCCGCTATCCTTGTGTATGGGGAT 900  
 Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
 Db 901 ATGCTCACTTTGATTGAAAGCTTCTCAGGAATACGATGACAAAGCCCTTGAATG 960  
 Qy 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340  
 Db 961 GCTGAGAAGATCGCGAAACACCTTGATTTTAAATGATTCATCGATCAATTAGATCAG 1020  
 Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 1021 GTCCAAATATGCGCGCATGCGAAGACTTGTCTCAAGATGATTCAGATATGGCCACAAT 1080  
 Qy 361 LysGlyLysLysLysMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
 Db 1081 CCAGCCCTTCAAAACATGAAGTGGATGAACGCCAGATTGCTCGTAAACGTCGCAATTG 1140  
 Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 TCTTCGATGACACCTGAGAGCGTGAACCCAGATTTGTTAAATCCAAAGCCGCTGCGCT 1200  
 Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 CGTATTGCTGCTGCTGCTGGAATACATTCGTCGAAGTCAATAAATTCATCAAGGACTTT 1260  
 Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440  
 Db 1261 AACCGGCTTAAACAGCTCATGCGGGTGTATGCTGGG----- 1299  
 Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454  
 Db 1300 -----GATATGAATAAATGATGAAGCAATGGGGATTAATCCAAATAACCTTCT 1350  
 RESULT 13  
 ID AAS55519  
 XX AAS55519 standard; DNA; 1572 BP.  
 AC AAS55519;  
 XX

DT 13-FEB-2002 (first entry)  
 XX Streptococcus pneumoniae DNA for cellular proliferation protein #90.  
 DE Antisense: ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX Streptococcus pneumoniae.  
 OS Streptococcus pneumoniae.  
 XX WO200170955-A2.  
 XX 27-SEP-2001.  
 XX 21-MAR-2001; 2001WO-US09180.  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253825P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 DR WPI; 2001-611495/70.  
 DR P-PSDB; AAU37660.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX Claim 27; Seq ID No 9156; 511pp; English.  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 1572 BP; 476 A; 300 C; 386 G; 410 T; 0 other;

Alignment Scores:  
 Pred. No.: 2.65e-109 Length: 1572  
 Score: 1307.00 Matches: 253  
 Percent Similarity: 74.73% Conservative: 90  
 Best Local Similarity: 55.12% Mismatches: 102  
 Query Match: 57.48% Indels: 14  
 DB: 23 Gaps: 2  
 US-09-943-108A-2 (1-455) x AAS55519 (1-1572)  
 QY 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCATTTTGAAGTTTAAACAGACGCTTTTGAAGACGCTTTTAAATAATCTACGTAAAAA 60  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetArgGlnValArgLeuAlaLeuPhe 40

Db 61 GGAATAATCTCTGAATCTGATCTCCAGAGCAACCAAGAAATTCGCTGSCCTGCTC 120  
 QY 41 GluAlaaspValasnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 121 GAGCGCAGCGTTCGCTTGTGTAAGACTTTATCAAGAAAGTTTCGTGACGCGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleValGln 80  
 Db 181 GTCGGGATGAGTGCATTCATCACTTAATCTCCGCAACAGATTATTAAATGTTGAT 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleasnMetSerasnLysPro 100  
 Db 241 GAGGAAGTACACCGCTTTAGTTCTCATAGCGCAGAAATTATCAAGTCACCTAAGATT 300  
 QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCAACCATCATGATGATGTTGTTTACAGGGCTGTGTAACACACCTTCTGTTGTTAAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrrasnLysLysProMetLeuValAlaAlaAspIle 140  
 Db 361 TTGGCCAACTCAAGAAAGAAATCTCTCTTGTGATGTTGCGCGGATATT 420  
 QY 141 TyrArgProAlaAlaIleasnGlnLeuGlnThrValGlyLysGlnIleaspIleProVal 160  
 Db 421 TATGTCACAGCTGCCATTGACAGCTTAAGACCTTGGACACAGATTGATGCTGCTC 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTTGCACCTGGACAGAAAGTACAGCTGTTGAGATTGTACGTCAAGTTTGGAGCAAGCC 540  
 QY 181 LysGluGluHisLeuaspPheValIleIleaspThrAlaGlyArgLeuHisIleaspGlu 200  
 Db 541 CAACTAATCATACAGACTATGCTTGTGATGATCTCGGGTCTGTTGAGATTGATGAG 600  
 QY 201 AlaLeuMetAsnGluLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 CTCCTCATGATGAGCTTCGTGATGTGAAGCATGTGCTCAACCAATGAATCTGCTT 660  
 QY 221 ValValaspSerMetThrGlyGlnaspAlaValasnValAlaGluSerPheaspGln 240  
 Db 661 GTCGTTGATGCTATGATGTTGTCAGGAAGCAGCACTGTTGCGGTGAGTTTAACTCTCAG 720  
 QY 241 LeuaspValThrGlyValThrLeuThrLysLeuaspGlyAspThrArgGlyGlyAlaAla 260  
 Db 721 TTGGAAGTACGTGGGGTCATCTTACAGATGATGCGGCTACTGCTGTTGCTGCTCT 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 Db 781 CTGCTGTTGCTCATTTACTTGGAAAAACCAATCAAGTTCACTGTGTACAGTGAAGAAT 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 ACGGACATTTGAACCTTCCACCGACGCGATGCTAGCGTATCTTGTGATGGGAT 900  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnaspValaspGlnGlnLysAlaLysaspLeu 320  
 Db 901 ATGCTCATTTGATGAGAAGCTTCTCAGGAATACAGTGAACAAAAGCCCTTGAATG 960  
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspaspPheLeuGluGlnLeuaspGln 340  
 Db 961 GCTGAGAAGATGCGGAAAAACCTTTGATTTTAAATGATTTTCATCGATCAATTAGATCAG 1020  
 QY 341 ValLysasnLeuGlyProLeuaspaspIleMetLysMetIleProGlyMetasnLysMet 360  
 Db 1021 GTGCAAAATATGGGCGCGATGGAAGACTTGCCTCAAGATGATTCAGGTATGCGCAACAAT 1080  
 QY 361 LysGlyLeuaspLysLeuasnMetSerGluLysGlnIleaspHisIleLysAlaIleIle 380  
 Db 1081 CAGCCCTTCAACATGAGAGGTGGATGAACCGCAGATTGCTGTAAACGTCGATTTGG 1140  
 QY 381 GlnSerMetThrProAlaGluArgasnAsnProAspThrLeuasnValSerArgLysLys 400



Db 841 ACGGACATTGAACCTTCACCCAGACCGCATGCTAGCGGTATCTTGGTATGGGGAT 900  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 Db 901 ATGCTCACTTGTATGAGAAAGCTTCTCAGGAATACGATGACAAAGACCCCTTGAATG 960  
 QY 321 GluLysLysMetArgGlnSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340  
 Db 961 GCTGAGAAGATGCGCGAARACACCTTGTATTTATGATTTCATGATCAATATGATCAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 1021 GTGCAAAATATGCGCGATGGAAGACTTGTCTCAAGATGATTCAGGTATGSCCAACAAT 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLe 380  
 Db 1081 CCAGCACTTCAAAACATGAAGTGGATGAACGCCAGATTCTCGTAAAGCTGCCATTGTG 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 TCTTCGATGACCTGAGAGACGTGAARACCCAGATTGTTTAATCCAGCCGTCGCGT 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 CGTATTGCTGCTGCTTCTGGAATACATTCTCGAAGTCAATAAATTTATCAAGGACTT 1260  
 QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysLys 440  
 Db 1261 AACCCAGGCTAACACGTCATGACAGGCTGTATGCTGGG----- 1299  
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454  
 Db 1300 -----GATATGATAAATGATGAGCAATGGGATTATCAATCAATCACTTCT 1350  
 RESULT 15  
 ID AAH90724  
 AC AAH90724 standard; DNA; 1572 BP.  
 XX AAH90724;  
 DT 02-OCT-2001 (first entry)  
 XX CFE 28 coding sequence.  
 KW Antibacterial; vaccine; gene therapy; bacterial cell wall viability;  
 KW CFE; CEG; Conserved Essential Gene; bacterial infection;  
 KW antisense therapy; antibiotic resistance; ds.  
 XX Streptococcus pneumoniae.  
 OS WO200149721-A2.  
 PN 12-JUL-2001.  
 XX 29-DEC-2000; 2000WO-US35604.  
 XX 30-DEC-1999; 99US-0174089.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;  
 XX Thanassi JA;  
 XX WPI; 2001-496721/54.  
 DR P-PSDB; AAM01025.  
 XX Nucleic acids encoding conserved essential genes involved in bacterial  
 PT replication which are potential targets for the treatment of antibiotic  
 PT resistant bacterial infections -  
 XX Claim 16; Page 200; 380pp; English.  
 XX The present invention relates to nucleic acids (AAH90701-AAH90918)

CC encoding polypeptides (AAM01002-AAM01114), which are essential for the  
 CC viability of a bacterial cell wall. The acronym CFE stands for "CFE For  
 CC Expression", where CFE stands for "Conserved Essential Gene". The nucleic  
 CC acids are useful for detecting the presence of proteins essential for the  
 CC viability of a bacterial cell wall in samples such as cells, tissues,  
 CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,  
 CC and for detecting corresponding target nucleic acid molecules with  
 CC complementary sequences. The nucleic acids are also useful for  
 CC determining whether a genomic nucleotide sequence of interest is  
 CC essential for viability of a bacterial cell or whether it resides within  
 CC an operon, by integrating an exogenous nucleotide sequence comprising a  
 CC portion of an open reading frame of the genomic sequence of interest  
 CC (comprising 200-500 base pairs) into the genomic sequence of interest  
 CC which confers a selectable phenotype to the cell, and determining cell  
 CC viability with a selection agent such as chloramphenicol. The nucleic  
 CC acids and proteins are also useful as vaccines and for treating bacterial  
 CC infections with gene therapy and antisense therapy. The nucleic acids  
 CC also enable identification of targets suitable for the treatment of  
 CC antibiotic resistant bacterial infections.  
 XX  
 SQ Sequence 1572 BP; 479 A; 298 C; 384 G; 411 T; 0 other;

Alignment Scores:  
 Pred. No.: 3,26e-109 Length: 1572  
 Score: 1306.00 Matches: 252  
 Percent Similarity: 74.73% Conservative: 91  
 Best Local Similarity: 54.90% Mismatches: 102  
 Query Match: 57.43% Indels: 14  
 DB: 22 Gaps: 2

US-09-943-108A-2 (1-455) x AAH90724 (1-1572)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCATTTGAAGCTTTAACAGAACGCTTTCAGAACGCTTTTAAATAATACGTAAAAA 60  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 61 GGAATAATCTCTGATCTGATGTCGAAGAGGCAACCAAGAATAATTCGCTTGGCTTGCTC 120  
 QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 121 GAGGCGCAGCTTGCTTGCCTGTGTAAAGACTTATCAAGAAAGTCTGCTGAGCGTGA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 Db 181 GTCGGCATGAGGTCATTGATACACTTAATCTCGCAACAGATTATTAATAATCGTTGAT 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 241 GAGGACTGACAGCGCTTTAGGTTCTGATACGCGAGAAATATCAAGTCACCTAAGATT 300  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCAACCATCATCATGATGATGTTGTTTACAAGGGCTGTGTAACAAACACCTTGTGCTGATA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyraAsnLysProMetLeuValAlaAlaAspIle 140  
 Db 361 TTGGCCCAACAACCTCAAGAAAGAAATAATGCTCTCTTTGATGATGCGCGGATATT 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 421 TATCGTCCAGCTGCCATTGACAGCTTAAGACTTGGGACACAGATTGATGCTGCTGTC 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTTCACCTTGAACAGAGAAGTACCAGCTGTTGATGATGTACGTCAAGGTTTGGAGCAAGCC 540  
 QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200  
 Db 541 CAACTAATCATCAACGACTATGCTTGTATGATGATGCTGCGGTGCTTGGCAGATTGATGAG 600  
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220

```

Db 601 CTCTCATGATGAGCTTCGTGATGTAAGACATTGGCTCAACCAATGAAATCTTGCTT 660
Qy 221 ValValaspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTCGTTGATGCTATGATTGTCAGAGACCAATGTCGCGTGTAAATGCTCAG 720
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 TTGGAAGTGACGGGTCTATCTTACCAGATGATGGGATACCTCTCGTGGTGGTCT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGTCTGTCTGCATCTACTGGAACCAATCAAGTTCAGTGTACAGGTGAAAAGATT 840
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACGGACATTGAACCTTCCACCCAGCCGATGCTAGCCGTATCCTTGGTATGGGGAT 900
Qy 301 ValLeuSerLeuIleGluLysAlaGlnInAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 ATGCTCACITTTGATTCAGAAAGCTTCTCAGGANTACGATGACAAAGCCCTTGAATG 960
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAGATCGCGGAAACACCTTTGATTTAATGATTCATCGATCAATTAGATCAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTCAAAATATGGGCCGATGGAGACTTCTCAAGATGATTCAGGTATGGCAACAAT 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCCCTTCAAACATGAAGTGGATGACGCCAGATTGCTGTAAACGTGCCATTGTG 1140
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACACCTGAAGAGCGTGAACCCAGATTGTTAAATCCCAAGCCGTCGCCGT 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTGCTGGTCTGGAAATFACATTCGCAAGTCAATAAATTCATCAAGGACTTT 1260
Qy 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
Db 1261 AACGAGCTAAACAGCTCATGCGGTGTTATGCTGGG----- 1299
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATGATGAAGCAAAATGGGATTAATCCAAATAACCTTCCT 1350

```

Search completed: February 25, 2003, 01:52:44  
Job time : 421 secs



```
Db 61 GGTAAACTTACGAGCTGATATAAGATAATGATGGCTGAAGTAAGATACGGTATT 120
Qy 41 GluAlaAspValAsnPhelLysValValLysGluPheLeuLysThrValSerGluArgAla 60
Db 121 GAGCGTGACGTAACCTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGGCA 180
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLeuValGln 80
Db 181 TTAGGTTCGGATGTAATGCATCATTAACACCGGCAACAAGTTATTAATAAGTTCAA 240
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLeuAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAAGTTGATGGGTGAGAAATATACATCGATTATATGTCATTAACCA 300
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTGTGTTATGATGGTTGGTTTACAAAGCTGCTGTAACCAACACACTGAGGTAAA 360
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATATTGATCGGTAATAATACAAACAAACCTATGTTAGTTCACGCGATATT 420
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATAATCAATCAACACAGTAGGGAACCAATTTGATATTCCTGTA 480
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAGCAGCACCAACAAATTTGAATCATGCAATTAACATGCT 540
Qy 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAGAGAACAATTAGACTTTGTAATCATTTGATACAGCAGTCGATACACATCGATGAA 600
Qy 201 AlaLeuMetAsnGluLeuLysGlnValLysGluIleAlaLysProAsnGluIleKetLeu 220
Db 601 GCATTGATGACGAATTAAGAGAGTAAAGAAATTTGTAACCAACGAAATTAATGTA 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTTGTCGATCATCATCGCGGTCAAGATGCTGTCATATGTCAGAACTTTTGACCATCAA 720
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTACCTTAACATAATTAGATGCTGATACACGTGCTGTCAGCT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTATCTATTCTGTCGTCGACACAAACCAATTAATTTGTTGTATGATGTAAGTA 840
Qy 281 AspGlyLeuGluPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGGTTAGACTATTCATCTGTAACGTATGCGCATCAGTATTTAGGTATGGTGAT 900
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTTAAGTTTAAATGAAAACCCACACAGATGTCGATCAGAAAGCAAGATTA 960
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAAGATCGGTGACTATCTGTTTACTTTAGATGATTTTTTATAGAACACTGATCAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATCTAGGACCACTGGATATTAATGATAAATGATTCAGGATTAATAAATG 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGGCTAGATAAGCTTAATATATGAGTGAAGCAAAATTTGATATTAAGCGCATATC 1140
Qy 381 GlnSerMetThrProAlaGluAsnAsnProAspThrLeuAsnValSerArgLysLys 400
```

```
Db 1141 CAGTCATGATGACCGGCTGAAGAAACAATCCAGACACACATTCATGATATCAGTAAAG 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTAAAGGCTGCTGCTGTTCAATTAACAAGAAAGTCAATCGTTTATGATGAAACAATT 1260
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
Db 1261 AACGATATGAAGAAATGATGAACAATTCACATGCTGGCGGTAAAGGTAAAGGTAAAGTAA 1320
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCATCAATGCAAAATATGTTAAAGGTATGATTTACCGTTT 1365
```

## RESULT 2

```
US-09-134-001C-1958
; Sequence 1958, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCI
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1958
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1958
```

```
Alignment Scores:
Pred. No.: 2,35e-165 Length: 1014
Score: 1534.00 Matches: 307
Percent Similarity: 97.29% Conservative: 16
Best Local Similarity: 92.47% Mismatches: 9
Query Match: 67.46% Indels: 0
DB: 4 Gaps: 0
```

US-09-943-108A-2 (1-455) x US-09-134-001C-1958 (1-1014)

```
Qy 1 MetalApheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 16 ATGGCATTTGAAGGATATCCGATCGCTTACAGCCACGATGCAAAAATCGGTGATAA 75
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 76 GGAAAGTAAACAGACAGACATATTAACAGATGATGCGTGAAGTGAAGATTACGTTATTG 135
Qy 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 136 GAAGCGGATGTTACTTCAAGTTGTTAAGGAATTTGTTAAGATGTTTCAGACGAGCG 195
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80
Db 196 CTAGGTTCTGATGTCATGCAATCTTTAAACCTGCGCAACAGGTTATTAATAATCGTACA 255
Qy 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 256 GAGAACCTTACTAGTTATGTTGGTGGAGAAATACTTCCATTAAGATGCGCAACAAACCA 315
Qy 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 316 CCAACTGTTGTCATGATGTCGCTTACAGTCCAGTAAACGACGACGACGAGTAAG 375
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 376 TTGGCATTAATTAATCGTAAAAAATAACAAAAAACCTTTACTTTGTCGACGAGATATT 435
```

QY	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
Db	436	TATCGTCCAGCTGCTATTGATCAATACAAACAGTAGGTAAACAAATTCATCCCTGTG	495
QY	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
Db	496	TATAGTGAAGGTGATCAAGTATCACACACAAATGTTGAATCGTCTTAAACATGCT	555
QY	181	LysGluGlnHisLeuAspPheValIleIleAspThrLacGlyArgLeuHisIleAspGlu	200
Db	556	AAAGAAACAATCTAGATTCGTTATCATTTATGATACAGCTGGTGTACACATTGATGAA	615
QY	201	AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu	220
Db	616	GCATTTATGAATGAGCTTCAAGAAGTTAAGAATCTCTAAACACAGCAAAATATGCTT	675
QY	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln	240
Db	676	GTTCGTGATGCAATGACAGGTCAGATGCTGTGAATGTGGCACAAATCAITTCGATGACAA	735
QY	241	LeuaspValThrGlyValThrLeuThrLysLeuaspGlyAspThrArgGlyAlaAla	260
Db	736	TTAGATGTTTTCAGGTGTAACATTCACATAATAGATGGTGATACACGCGTGTGCAACA	795
QY	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
Db	796	CTTTCATCGTTCGTTACCCAAAACCTATTAAATTTGTAGTATGAGTCAGAAATG	855
QY	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
Db	856	GATGGTTTAGAATATTCCATCCTGACCAATGGCGTCCAGTATTTTAGTATGGTGAT	915
QY	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValaspGlnGluLysAlaLysAspLeu	320
Db	916	GTTTTRAGCTCATTTGAATAAGCCACACAGATGTAGATCAGAAAGGCTAAAGATTAA	975
QY	321	GluLysLysMetArgGluSerSerPheThrLeuAsp	332
Db	976	GAGAAGAAGATGAGAGAGTCATCTTCCTCTCTAAAT	1011
RESULT 3			
	US-08-961-527-46/c		
	; Sequence 46, Application US/08961527		
	; Patent No. 6420135		
	GENERAL INFORMATION:		
	; APPLICANT: Charles Kunsch		
	; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences		
	; NUMBER OF SEQUENCES: 391		
	; CORRESPONDENCE ADDRESS:		
	; ADDRESSEE: Human Genome Sciences, Inc.		
	; STREET: 9410 Key West Avenue		
	; CITY: Rockville		
	; STATE: Maryland		
	; COUNTRY: USA		
	; ZIP: 20850		
	COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage		
	; COMPUTER: HP Vectra 486/33		
	; OPERATING SYSTEM: MSDOS version 6.2		
	; SOFTWARE: ASCII Text		
	; CURRENT APPLICATION DATA:		
	; APPLICATION NUMBER: US/08/961,527		
	; FILING DATE:		
	; CLASSIFICATION: 424		
	; PRIOR APPLICATION DATA:		
	; APPLICATION NUMBER:		
	; FILING DATE:		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Brookes, A. Anders		
	; REGISTRATION NUMBER: 36,373		
	; REFERENCE/DOCKET NUMBER: PB340P1		
	; TELECOMMUNICATION INFORMATION:		



```

|||||:||||| :||| ||||||||| || | |||||:
2253 CTGCTGTTCGTGCATTAAGTGAACCAACTCAAGTTCACTGGTAGCGTGAAAGATT 2194
QY 281 Asp6gLYLeuGlnLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
:|||||:||||| :||| ||||||||| || | |||||:
2193 ACGSACAATGAACCTCCACCACGACGCGATCTAGCGGTATCCTTGGTATGGGGAT 2134
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
:|||||:||||| :||| ||||||||| || | |||||:
2133 ATGCTCACTTGGATTGAGAAGCTTCTCAGGAATACGATGAACAAAAGCCCTTGAATG 2074
QY 321 GluLYLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
:|||||:||||| :||| ||||||||| || | |||||:
2073 GCTGAGAAGATGCCGAAACACCTTGAATTAAATGATTTTCATCGATCAATTAGATCAG 2014
QY 341 ValLYsAsnLeuglyProLeuaspAspIleMetLYsMetIleProGlyMetAsnLYsMet 360
:|||||:||||| :||| ||||||||| || | |||||:
2013 GTGCAAAATATGGGCGCATGGAAGACTTGTCAAGATGATTCAGGTATGCCAACAAAT 1954
QY 361 LysGLyLeuAspLYsLeuAsnMetSerGluLYsgLnIleAspHisIleLYsAlaIlelle 380
:|||||:||||| :||| ||||||||| || | |||||:
1953 CCAGCCCTTCAAACATGAAGTGGATGAAGCCGCAATTCGTCGTAAACGTGCCAATTGTG 1894
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLYsLYs 400
:|||||:||||| :||| ||||||||| || | |||||:
1893 TCITCGATCACACCTGAAGAGCGTGAAACCCAGATTCTTAAATCCAACGCGTCGCGGT 1834
QY 401 ArgIlealALysLYsSerGlyrArgSerLeuGlnGlnValAsnArgLeuMetLYsGlnPhe 420
:|||||:||||| :||| ||||||||| || | |||||:
1833 CGTATTTCGTGCTGGTCTTGGAATAATCATCTCGAAGTCAATAATTCATCAAGGACTTT 1774
QY 421 AsnAspMetLYsLYsMetMetLYsGlnPheThrGLyLYsGlyLYsLYsGlyLYs 440
:|||||:||||| :||| ||||||||| || | |||||:
1773 AACAGGCTAAACAGCTCATCGAGGCTGTATGTCTGGG----- 1735
QY 441 ArgAsnGlnMetGlnAsnMetLeuLYsGlyMet-----AsnLeuPro 454
:|||||:||||| :||| ||||||||| || | |||||:
1734 -----GATATGAATAAAATGATGAAGCAAAATGGGATTAATCCAAATAACCTTCT 1684

RESULT 4
US-09-035-382-3
; Sequence 3, Application US/09035382
; Patent NO. 6284515
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES
; TITLE OF INVENTION: AND POLYNUCLEOTIDES
; FILE REFERENCE: GM50035
; CURRENT APPLICATION NUMBER: US/09/035.382
; CURRENT FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: 60/057.890
; EARLIER FILING DATE: 1997-09-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-035-382-3
```

```

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgLuuValargLeuAlaLeuPhe 40
    |||||
Db 61 GSTAAACTTACTGAAGCTGATATAAAGATAATGATGCGTGAAGTAAAGATTAGCGTTATT 120
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
    |||||
Db 121 GAGCTCAGCTAAACTTTAAAGTGGTAAGAAGAAATTATTAAACAGTATCAAGACGGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
    |||||
Db 181 TTAGGTTTCOGATGTAATCAATCATTAACACAGGCAACAAGCTTATTAATAAGTTCAA 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
    |||||
Db 241 GATGAATTAACGAAGTTGATGGGTGAGAGAAAATACATCGATTAATATGTCAAAATAAACCA 300
QY 101 ProThrValValMetMetValClyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
    |||||
Db 301 CCTACTCTGTTGATGATGGTTTACAGGCTGCTGGTAAACAACAACACTGCAGGTAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
    |||||
Db 361 TTAGCATTAATGATGCGTAAAAAATACAAAAAACCTATGTAGTTGCAGCAGATATT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
    |||||
Db 421 TATGCTCCAGCAGCGTAATCAATACAAACAGTAGGGAACAATTCATATTCATCTCTGTA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
    |||||
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACCAACAAATTTGTACTAATGCAATTAACATGCT 540
QY 181 LysGluGlnHisIleuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspL 200
    |||||
Db 541 AAAGAAGAACATTTAGACTTTGTATTCATGATGATACGCGGTGGATACCAATCGATGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
    |||||
Db 601 GCATGTGATGAACGAATTAAGAAGTAAGAAAGAAATTTGCTAAACCAACGAATTTATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
    |||||
Db 661 GTTGCGATTCATGACGGGTCAAGATGCTGTCAATGTTGCAGAATCTTTTGACGATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
    |||||
Db 721 CTTGATGTCACAGGTGTTACCTTAACCTAAATAGATGGTGATACCCGTGTGTGCGAGCT 780
QY 261 LeuSerIleArg 264
    |||||
Db 781 TTATCTATTCTGT 792

RESULT 5
US-08-923-772-1
: Sequence 1, Application US/08923772
: Patent No. 5972651
: GENERAL INFORMATION:
: APPLICANT: Black, Michael T.
: TITLE OF INVENTION: NOVEL IFI
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert Price & Rhoads
: STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0

```

; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/923,772  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Dickinson, Todd Q  
 ; REGISTRATION NUMBER: 28,354  
 ; REFERENCE/DOCKET NUMBER: GML0080  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-994-2252  
 ; TELEFAX: 215-994-2252  
 ; TELEX:

; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1569 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-923-772-1

Alignment Scores:  
 Pred. No.: 4,49e-138 Length: 1569  
 Score: 1297.00 Matches: 251  
 Percent Similarity: 74.5% Conservative: 91  
 Best Local Similarity: 54.68% Mismatches: 103  
 Query Match: 57.04% Indels: 14  
 DB: 2 Gaps: 2

US-09-943-108a-2 (1-455) x US-08-923-772-1 (1-1569)

QY 1 MetAlapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCATTTGAAGCTTTAAACAGAACGTTTGAGAACGCTTTTAAATAATCTACGTAATAAAA 60  
 QY 21 GlyLysLeuThrGluAlaAspLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
 Db 61 GGAATAATCTGATCTGATGTCGACGAGGACGACGACGACGACGACGACGACGACGACGAC 120  
 QY 41 GluAlaAspValAsnPhelLysValLysGluPhelLysThrValSerGluArgAla 60  
 Db 121 GAGCGCGAGTGTGCTTGGCTTGTGTAAGGACCTTTATCAAGAAAGTGTGTCGAGGTCGA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysLysLys 80  
 Db 181 GTCGGGCGATGAGTGTGATGATCACTTAATCTGCGCAACAGAGATTATTAATACTGTTGAT 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysAsnMetSerAsnLysPro 100  
 Db 241 GAGGACCTGACAGCGTTTGTAGTCTGATACGCGCAGAAATATCAAGTCACCTAAGATT 300  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCAACCATCATCATGATGTTGTTTACAAAGGGGTGTGTAACAAACCTTTGCTGGTAAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspLys 140  
 Db 361 TTGGCCACAACTCAGAAAGAGAAAGAAATGCTGCTCTTGTGATTCGGCGGATATT 420  
 QY 141 TyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLysLysLysLysLys 160  
 Db 421 TATGCTCCAGCTGCCATGACCAAGCTTAAGACCTTGGGACACAGATTTGATGCTGCTC 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLysLysLysLysLysLysLysLysLys 180  
 Db 481 TTTGCACTTGGCAACAGAAAGTACCAGCTGTTGATTTGATTTGATTTGATTTGATTTGAT 540  
 QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisLysLysLys 200  
 Db 541 CAAACTAATCATACGACTATGCTTGTGATGATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCT 600

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluLysGluLysGluLysGluLysGluLys 220  
 Db 601 CTCCTCATGAATGAGCTTCGATGTGAAAGTATTGGCTCAACCAAAATGAATCTTGGCTT 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAspAlaGluSerPheAspAspGln 240  
 Db 661 GTCGTTGATGCTATGATGTTGTCAGAGACGACGACGATGTCGCGTGAGTTAATGCTCAG 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 721 TTGAAGTGAAGTGGGCTCATCTTACCAAGATTGATGATGATGATGATGATGATGATGAT 780  
 QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLys 280  
 Db 781 CTGCTCTTCGTCATCATCACTGGAACCAATCAAGTTCACGTGACAGTGAAGTGAAGTGA 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 ACAGATATCGAAACCTTCCACCAGACGCTATGCTAGCCGATGCTTGGCATGGGGAT 900  
 QY 301 ValLeuSerLeuLysGlnAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
 Db 901 ATGCTCACTTGTGATGAGAAAGCTTCTCAGGAATACGATGAACAAAGAGCCCTTGAATG 960  
 QY 321 GlnLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340  
 Db 961 GCTGAGAGATGCGCGAAACACCTTTGATTTAATGATTTAATGATTTAATGATTTAATG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspLysMetLysMetLysMetLysMetLysMetLysMet 360  
 Db 1021 GTGCATAAATATGGGCGGATGGAAGCTTGTCTCAAGATGATTCAGGATGATTCAGG 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGlnLysGlnLysGlnLysGlnLysLysLys 380  
 Db 1081 CCAGCATCTCAAAACATCAAGGTGATGAGCGGACGATGCTCGTAAACGGTCCGCTTGG 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 TCTTCGATGACATCTCAAGACGTCGAAACCCAGATTTGTTAAATCCAGACCGCTCGCG 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 GGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 QY 421 AsnAspMetLysLysMetLysLysLysLysLysLysLysLysLysLysLysLysLys 440  
 Db 1261 ACCAGGCTAAACGCTCAAGCGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1299  
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454  
 Db 1300 -----GATATGAATAAATGATGAAGCAATGGGATTAATCCAAATAACCTTCT 1350

# RESULT 6

US-09-385-287-1  
 ; Sequence 1, Application US/09385287  
 ; Patent No. 6350857  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Black, Michael T.  
 ; TITLE OF INVENTION: NOVEL fih  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dechert Price & Rhoads  
 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: US  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/09/385,287
; FILING DATE: 30-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/923,772
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1569 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-385-287-1
Alignment Scores:
Pred. No.: 4,49e-138 Length: 1569
Score: 1297.00 Matches: 251
Percent Similarity: 74.51% Conservative: 91
Best Local Similarity: 54.68% Mismatches: 103
Query Match: 57.04% Indels: 14
DB: 4 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-385-287-1 (1-1569)
Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGCGATTGGAAGTTTAAACAGACGTTTGCAGAGCGCTTTAAAAATCTACGTAAAAA 60
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAATAATCTGATCTGATGTCAGAGGCAACCAAGAAATGCGTTGGCTTGCTC 120
Qy 41 GluAlaAspValAspPheLysValValLysGluPheLleLysThrValSerGluArgAla 60
Db 121 GAGCGCGAGCTTGCTTCCTCCCTTGTAAAGGACTTTATCAAGAAGTTCGTGAGCGTGA 180
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTGGGCGATGAGTCATTGATACATTAACTCTGCCACACAGATTATTAATTCGTTGAT 240
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAGGAATGACAGCGCTTTAGGTCTGTATACGCGCAAGAAATATCAAGTCACTAAGATT 300
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCAACCATCATGATGTTGGTTTCAAGGGCGTGGTAAACAAACCTTTGCTGTGATAA 360
Qy 121 LeuAlaLeuLeuMetArgLysLysTyraAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTGGCCACAACATCAAGAAGAGAAATGCTGCTCCTTTGATGATGCGGCGGATATT 420
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCTGCCATTGACCATTAAGACCTTGGGACACACAGATTGATGTCGCTGC 480
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTTCGACTTGGACACAGATACCAAGCTGTTGAGATTGATGCTCAAGGTTGGAGCAAGCC 540
Qy 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 CAACACTAATCAACAGCATGCTTCTGATTGATGCTGCGGGTCTGTTGCAGATTGATGAG 600

```

```

Qy 201 AlaLeuMetAsnGluLeuLysGlnValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 CTCCTCATGATGAGCTTCGTGAITGAAGATATTGGCTCAACCAATGAAATCTGCTT 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTCGTTGATGCTATGATTGCTCAGGAAGCAGCAATGTTGGCGTGAGTTAATCTCTCAG 720
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAla 260
Db 721 TTGGAAGTACTGGGGTCATCCTTACCAAGATTGATGCTGATATCTGCTGCTGCTCT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGCTGTTGCTCACATCACTGGAAAACCAATCAAGTTCACTGGTACAGTGAAAAATT 840
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetalaserArgIleLeuGlyMetGlyAsp 300
Db 841 ACAGATATCGAAACCTTCCACCAGACCGTATGCTAGCCGTATCTTGGCATGGGGAT 900
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 ATGCTCACTTTGATTGAGAAGCTTCTCAGGAATACGATGAACAAAAAGCCCTTGAATG 960
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
Db 961 GCTGAGAGATCGCGGAAACACCTTTGATTGTTTAAATGATTCATCGATCAATAGATCAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTCAAAAATATGGGCGGATGGAAGACTTGTCAAGATGATTCAGGTATGCGCAACAAT 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
Db 1081 CCAGCACTTCAAAACATCAAGGTGGATGAACCCAGATGCTGTAACGTGCGCATTTGTG 1140
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTGATGATGACATCTGAAGAACGTGAAACCCAGATTTGTAAATCCAAAGCCGTCGCGT 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CPTATTGCTGCTGTTGTTGGAATACATTCGTCGAAGTCATTAATATCAAGGACTTT 1260
Qy 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
Db 1261 AACAGGCTAAACAGCTCATGCGAGGTGTTATGTCGGG----- 1299
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAAATGATGAAGCAATGGGGATTAATCCAAATAACCTCTCT 1350

```

## RESULT 7

US-09-221-017B-928/c

; Sequence 928, Application US/09221017B

; Patent No. 6444799

; GENERAL INFORMATION:

; APPLICANT: Ross, Bruce C.

; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON &amp; FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

```

1 / APPLICATION NUMBER: US/09/221,017B
2 / FILING DATE: 23-DEC-1998
3 / CLASSIFICATION:
4 /
5 / PRIOR APPLICATION DATA:
6 /
7 / APPLICATION NUMBER: PP1182
8 / FILING DATE: 31-DEC-1997
9 /
10 / PRIOR APPLICATION DATA:
11 /
12 / APPLICATION NUMBER: PP1546
13 / FILING DATE: 30-JAN-1998
14 /
15 / PRIOR APPLICATION DATA:
16 /
17 / APPLICATION NUMBER: PP2511
18 / FILING DATE: 09-APR-1998
19 /
20 / PRIOR APPLICATION DATA:
21 /
22 / APPLICATION NUMBER: PCT/AU98/01023
23 / FILING DATE: 10-DEC-1998
24 /
25 / ATTORNEY/AGENT INFORMATION:
26 /
27 / NAME: MONROY, Gladys H
28 /
29 / REGISTRATION NUMBER: 32,430
30 /
31 / REFERENCE/DOCKET NUMBER: 27340-2000
32 /
33 / TELECOMMUNICATION INFORMATION:
34 /
35 / TELEPHONE: 650-813-5600
36 / TELEFAX: 650-494-0792
37 /
38 / TELEX: 706141
39 /
40 / INFORMATION FOR SEQ ID NO: 928:
41 /
42 / SEQUENCE CHARACTERISTICS:
43 /
44 / LENGTH: 1098 base pairs
45 /
46 / TYPE: nucleic acid
47 /
48 / STRANDEDNESS: double
49 /
50 / TOPOLOGY: circular
51 /
52 / MOLECULE TYPE: DNA (genomic)
53 /
54 / HYPOTHETICAL: NO
55 /
56 / ANTI-SENSE: UNKNOWN
57 /
58 / ORIGINAL SOURCE:
59 /
60 / ORGANISM: PORYPHYROMONAS GINGIVALI
61 /
62 /
63 / FEATURE:
64 /
65 / NAME/KEY: misc_feature
66 / LOCATION: 1...1098
67 /
68 / US-09-221-017B-928

```

[illegible]

QY 137 AlaAlaAspIleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGluIle 156  
Db 138 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 741 GCCTCGCAGGTGTATCGCCCTCGGGGTATTCACAGACAGTGATGATCCTCGGAGACAGCTC 682

QY 157 AspIleProValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAla 176  
Db 158 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 681 GGGGTACCCTGTACAGCAACCGGACACCAAAGAAGCGGTGAGATAGACCCGCACAGCC 622

QY 177 LeuLysHisAlaLysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeu 196  
Db 178 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 621 ATACGCCAAGCAACCAAGGGAGCAGACTGTCATCATCGACACGGCGGTGCTTTG 562

QY 197 HisIleAspCpuAlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsn 216  
Db 198 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 561 GCCATTCCAGCAAGAAATGATCGCGAGATAGAGGCCATCAAGGCTCCATCCACCGCAAC 502

QY 217 GluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSer 236  
Db 218 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 501 GAAGTCCTTTTCGTGGTAGCTCGATGACCGGTCAAGGACCGCTGAATACACCAAGGAG 442

QY 237 PheAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspCylAspThrArg 256  
Db 238 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 441 TTCATAACACGCTTGATTCGACGGTGTTCCTTACCAAGCTCGATGGAGATACCCGG 382

QY 257 GlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMet 276  
Db 258 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 381 GGCGGTGCCGCCCTCGATCCGTGCTGTAAACAAGCCCATCAAGTCTGCGGTACG 322

QY 277 SerGlnLysLeuAspCylLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeu 296  
Db 278 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 321 GGGGAGAGATGAAGCAATCGATGCTTTTCACCCCGAGCGTATGGCGACGATCCTT 262

QY 297 GlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnGlnAspValAspClnGluLys 316  
Db 298 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 261 GCATGGGGACATCGTATCCCTGTGTGGAGCTGCTCAGNGCAATACGACAGCGGAA 202

QY 317 AlaLysAspLeuGlnLysLysMetArgGluSerPheThrLeuAspPheLeuGlu 336  
Db 318 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 201 GCTGCAAACTGGAGAGAAAATACCAAGAATCAGTTCGACTTCAATGACTTCCTGCC 142

QY 337 GlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspileMetLysMetIleProgly 356  
Db 338 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 141 CAGATACATCAGATCAAAAGATGGGTACCTCAAGAGCTTCATCCATGATCCCGGA 82

QY 357 MetAsnLysMetLysGlyLeuAspLysLeuAsnMetSerGlnLysGlnIleAspHisIle 376  
Db 358 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 81 GTGGGA-----AAGGCCATCAAGGATATCGCATAGACGACGAGTCCCTTAAGAGCATC 28

QY 377 LysAlaIleIleGlnSerMetThrPro 395  
Db 378 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 27 GAGGCTATCATATTTCCATGACGCC 1

RESULT 8  
US-09-066-047-6  
; Sequence 6, Application US/09066047A  
; Patent No. 6506394  
; GENERAL INFORMATION:  
; APPLICANT: MORPHY, Cheryl  
; STOREY, James  
; BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF  
; ;  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,047A
; FILING DATE: 24-Apr-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/044,869
; FILING DATE: 25-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Superko, Colleen
; REGISTRATION NUMBER: 39,850
; REFERENCE/DOCKET NUMBER: 106,941,156
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-066-047-6

```

## Alignment Scores:

```

Pred. No.:          9,27e-87          Length:          4804
Score:              853.00             Matches:          174
Percent Similarity: 63.57%             Conservative:    100
Best Local Similarity: 40.37%           Mismatches:      147
Query Match:        37.51%             Indels:          10
DB:                  4                 Gaps:             3

```

US-09-943-108a-2 (1-455) x US-09-066-047-6 (1-4804)

```

QY   1  MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  3490  CTATGTTTAACTCTTTAACCAAGGGGTTTCTCTCGCGGCAAGAGTTAAGTGGRAAG 3549
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY   21  GlyLysLeuThrGluAlaAspIleLysIleMetMetArgLeuValArgLeuAlaLeuPhe 40
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  3550  CGGAGATATCCAGAGGATTTTGATCTGTGTAAGAGATATAACTCAGGCGATTGTG 3609
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY   41  GluAlaAspValAsnPheLysValLysGluPheLleLysThrValSerGluArgala 60
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  3610  GATCGGATGTAATCTGTGTGTGACGAATTTATAGAGACGTATAAAAGCAAGATC 3669
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY   61  LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLleLysIleValGln 80
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  3670  GTAGGGCGGATGATTAAGGGGTGTCGCCGAGCAATGTCATAAGCGTATAGAA 3729
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY   81  AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLeaSnMetSerAsnLysPro 100
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  3730  GAGTGTTCAGTGAAGTTAGTAATAGAGAGAGCGCTCTTGATCTTAAGGGAAGATT 3789
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY   101  ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  3790  CTTGCAATATCATGATGGTGGGCTTCAGGTGTGGTGAAGTACTTAACACAGTAAAA 3849
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY   121  LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLysValAlaAlaAspIle 140
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  3850  GTTGCACTGAGGTTA---AAGAGGATTCTAAACCCGTGGTAGCGCTTTAGACGTA 3906
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY   141  TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  3907  TATCTGCTCGAGCTCAGAACACAGTGAAGGTTTGGCTGATGAGTGTGTATACACAGT 3966
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY   161  TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuHisala 180
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

Db  3967  CTTCCATCTGTTGAGGAGCAAAAACCACTTGATATTGCGAAGCTGCTATGAGGGAAGC 4026
QY   181  LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4027  AGGCTCAAGGGCCAGCATGTGGTCTTTGGATACAGCGGGCGCTTGCATATCAATCAG 4086
QY   201  AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLys 220
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4087  GACATGATAGATGAGCTGAAGTGTGTAAAGAGAGGAGGTATCCACAGCTGAATTTGATTG 4146
QY   221  ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4147  GTGTAGACTCTTAAATGGGCAAGATGCCGTCACTAGTGGCAAGTTCAATGAGGAG 4206
QY   241  LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4207  TTAGGCATTACTGGAGCATCTTTACAGGGCGGATGGTGATCCTAGGGGTGGTCTATC 4266
QY   261  LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4267  TTGTCTATGAAGTTGTTGCTGGATGCTCTATAAAGTTCATGCTACGGGAGAGAGCCT 4326
QY   281  AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4327  GAAGATTTGGACGATTTCTATCTGATAGATAGCTCGTGAATGTTAAATATGGGAGAT 4386
QY   301  ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4387  GTGCGATCTCTGTAGAAAGCGGTAGAGCGGTTGGCAAGGATACAATTAATGAGCTA 4446
QY   321  GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4447  CAGGCGAAGGCCAAGAGGTAATAATTCGATTTGGATGATCTCTGTTATTCAGCTGAAGCT 4506
QY   341  ValLysAsnLeuGlyProLeuAspAspIleMetLysMetLysMetLysMetLysMetLys 357
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4507  TTGAATAAAATGGGTGATTTGCTAATAATAAGAGTTATACCCGCTTCGGTAACGAT 4566
QY   358  -----AsnLysMetLysLysLeuAspLysLeuAsnMetSerGluLysGlnIleAspHis 375
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4567  ATAAACCAAGTTGGCGGATA-----GCTGATGACAGCAAGTCGACATG 4614
QY   376  IleLysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsn 395
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4615  TACATTGCGATTATTAATCAATGACGAGAGGAGGCGAATCCTGAGATACTGAAT 4674
QY   396  ValSerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArg 415
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4675  GTGCGAGGAAGGCAAGATAGCAAGGTTGCGGAGTTAAGTTGATGCTGTAATCGC 4734
QY   416  LeuMetLysGlnPheAsnAspMetLysLysMet 426
    :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  4735  TTGCTAAAGCAGTATATATCAGATGAATTCGATA 4767

```

## RESULT 9

```

US-08-317-401E-3
; Sequence 3, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; APPLICANT: raver, Debbie Sue
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,401E
FILING DATE: 03-October-1994
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4248,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2166 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
/PS-08-317-401R-3

```

Alignment Scores:	
Pred. No.:	1.24e-57
Score:	592.50
Percent Similarity:	50.11%
Best Local Similarity:	33.04%
Query Match:	26.06%
DB:	2
Indels:	19
Gaps:	9
Matches:	156
Length:	2166

US-09-943-108A-2 (1-455) x US-08-317-401E-3 (1-2166)

QY	1	MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys	20
Db	86	ATGTGCTCTTCAGGATCTCGGCGCGCAATCAACGCGCGCTCAATGACTGACTCGCTCC	145
QY	21	GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
Db	146	AACAAITTTGACGAGAAGGCTTTGATGACATGATTTAAGAGATCTGCGCGCCCTCTCTG	205
QY	41	GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla	60
Db	206	TCGCGCGAGCTCAACGTCGCGCTGCTCCAGTCCCTCCGCAAGTCATCAANTCAGCGTC	265
QY	61	LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln	80
Db	266	AACTTTTGCTCTCTCTCCCTCGCGCGTGAACAAGAAGGTTTGATTCAAAGGCGCTTTC	325
QY	81	AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
Db	326	GATGAGCTGTTTCCCTCGTTTGATCCCATCGCGAGCCCTTCGCGCCCAAGAAGGGCGCC	385
QY	101	ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys	120
Db	386	TCCAACTGATCATGTTCGTGCGTTCGAGGTCGCGGTAACACCACTTGTACCAAG	445
QY	121	LeuAlaLeuLeuMetArgLysLysTyrAsnLys-----LysProMetLeuValAla	137
Db	446	CTGGCC-----CGCCACTACAGATGCGCGGCTTCAAGACTGCCTCGTGTGT	493
QY	138	AlaAspIleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGluIleAsp	157
Db	494	GCGGATACCTTCGCTGCTGTCCTTCGCACCACTGGAAGCAGAATGCCACCAAGCCAAAG	553
QY	158	IleProValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeu	177
Db	554	ATCCCTACTACGGTAGCTTCACCAACACGACCGCCCATCTGACACCGAGGTGTG	613
QY	178	LysHisAlaLysGluGlnHisLeuAspPheValIleLeuAspThrIleValArgLeuHis	197
Db	614	GCCAAAGTTCAAGAAGACGCTTTCGAATCATCATCATCGTTCGATACCAAGTGGTCAAG	673

Qy	198	lLeaspGluAlaLeuMetAsnGluLeuLysGluValLysGluAlaLysProAsnGlu	217
Db	674	CAGAAAGAGAGCTTTTCCACGAAATGACCCAGATATCAAGACCGCGTCCACCCCGACGAC	733
Qy	218	lLeMetLeuValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPhe	237
Db	734	ACCATCTGCTGCTCGACGAGCACACFCGGTCAAGCTGCGGAAAGCCAGTCTCTGCGCTTC	793
Qy	238	AspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGly	257
Db	794	AGGCCACCGCAGACTCTGGAGCCATCATCATCACCACGAGTGTCTACGCCGCGAGST	853
Qy	258	GlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSer	277
Db	854	GTTGTTCTATTTCGCGCTCGCGCCACACACATCCCATATTACCTCGGTACCGGT	913
Qy	278	GlnLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGly	297
Db	914	GAGCACTGATGGATCTGGAGACCTTTGAGCGGAGAGCGCTTCATCCAGAGACTCTGGT	973
Qy	298	MetGlyAspValLeuSerLeuIleGlnLysAlaGlnGln---AspValAspGlnGlnLys	316
Db	974	ATGGCGCATATGCTGCTGGTAGAGCAGCTACAAGCCGTGACCAAGGACTCTGCTCC	1033
Qy	317	AlaLysAspLeuGlnLysLysMetArgGlnSerSerPheThrLeuAspAspPheLeuGln	336
Db	1034	GCCAGAAACCTACAAGCACATCTCCGAAGGTATCTACACGCTCGGTGACTTCCGCGAG	1093
Qy	337	GlnLeuAspGlnValLysAsnLeuGlyProLeuAspPleuMetLysMetIleProGly	356
Db	1094	AACATCCTCATCATCAAGATGGAGCCCTCTCCAGCTCTCGGCATGATCCCGGT	1153
Qy	357	Met---AsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHis	375
Db	1154	CTATCCAACTGACCGCGGTCTCGATGACGAGACGCGCTCCATGAAG---CTCCGTGCG	1210
Qy	376	lLeuLysAlaIleGlnSerMetThrProAlaGluArgAsnProAspThrLeuAsn	395
Db	1211	ATGATCTACATCTCGACATGATGCGCGCGCGCACTCGATGGC---GATGGCAGAGAT	1267
Qy	396	Val-----SerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGln	411
Db	1268	TTGTCGAACACCCAGCGCGATGCTCGATGCTGCGGAGCGGTACACCGCTCGCG	1327
Qy	412	GluValAsnArgLeuMetLysGlnPheAsnAspMetLysLysMetLysGlnPheThr	431
Db	1328	GAAGTGAAGACGCTCTCTCCAGCACCGCATGATGCGCGCATGGCCCAAGCGTGC---	1384
Qy	432	GlyGlyGlyLysGlyLysLysLysLysArgAsnGlnMetGlnAsnMetLeuLysGlyMet	451
Db	1385	---GTTGACACAGACAGATGCGAGCGC-----GCCAGACATGCTCAAGGCGGT	1435
Qy	452	Asn	452
Db	1436	AAC	1438
RESULT 10			
US-08-317-401E-1			
; Sequence 1, Application US/08317401E			
; Patent No. 5922561			
; GENERAL INFORMATION:			
; APPLICANT: Thompson, Sheryl Ann			
; APPLICANT: Iyer, Debbie Sue			
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF			
; TITLE OF INVENTION: ASPERGILLUS NIGER			
; NUMBER OF SEQUENCES: 10			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: No. 59225610 No. 5922561disk of No. 5922561th America, I			
; STREET: 405 Lexington Avenue, Suite 6400			
; CITY: New York			
; STATE: New York			
; COUNTRY: U.S.A.			



```
; SEQ ID NO 20
; LENGTH: 4370
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-08-981-527A-20

Alignment Scores:
Pred. No.: 6,96e-39 Length: 4370
Score: 432.00 Matches: 114
Percent Similarity: 53.25% Conservative: 66
Best Local Similarity: 33.73% Mismatches: 120
Query Match: 19.00% Indels: 38
DB: 10 Gaps: 10

US-09-943-108a-2 (1-455) x US-08-981-527A-20 (1-4370)

QY 4 GluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeu 23
Db 2351 GATTCCGTATCTGAAAGTTTAAAGTATGCGCTTGAAGAAACAAAGAACTCCCTTCAAAAC 2410
QY 24 ThrGluAlaAspIleLysIleMetArgGluValArgLeuAlaLeuPheGlu----- 41
Db 2411 AAAGTGAATGATCTGTATCCCGTTACCGTAAAGTGGATGAGGATTCCTCGAAGAGCTT 2470
QY 42 -----AlaAspValAsnPhelLysValLysValLysGluPheIle----- 53
Db 2471 GAAGAGGTTCTTATCAGCGGGATGTCGGTTTACACCGCTTATGAATATATAGATGAG 2530
QY 54 -----LysThrValSerGluArgAlaLeuGlySerAspValMetGlnSerLeuThrPro 71
Db 2531 CTGAAAAAAGAGTCAAAACGAGA-----AATATTCAAGAT 2566
QY 72 GlyGlnGlnValIleLysIleValGlnAspGluLeuThrLysLeu-----MetGlyGly 89
Db 2567 CCAGAGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2626
QY 90 GluAsnThrSer-----IleAsnMetSerAsnLysProThrValValMetMetValGly 108
Db 2627 GAGCAAAATTCAGAACTGAACATCCAGGATGGCGCTTAAACGTAACTCTCTGTGTAGGT 2686
QY 109 LeuGlnGlnAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArgLysLys 128
Db 2687 GTAAACGGCGTCCGGAACAAACACAGATCCGAAAGCTGTCTATAAAATG---AAACAA 2743
QY 129 TyrAsnLysLysProMetLeuValAlaAlaAspIleThrArgProAlaAlaIleAsnGln 148
Db 2744 GAAGGAAATCTGTGTACTTGCCTGCGGAGACACTTTTAGCGCGGACCAATTAACAG 2803
QY 149 LeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGlyAspGlnValLys 168
Db 2804 CTGGAGATATGGGAGAGCTTACAGGAGTCTGTCTATTAAGCAGCAGCGGAGAGCGCAT 2863
QY 169 ProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluLysIleAspPheVal 188
Db 2864 CCGCGCGCTGTCTACGATGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 2923
QY 189 IleIleAspThrAlaGlyArgLeuHisIleAspGlnAlaLeuMetAsnGlnLysGlu 208
Db 2924 ATTTGTGTATGAGCGGAGCGCTCTCCAAACAAAGTAAATCTCATGAAGAGGTTTGAAGA 2983
QY 209 ValLysGluIleAlaLys-----ProAsnGluIleMetLeuValVal 222
Db 2984 GTAAACGTTGTATCGAAGAGAGAGTCTCTGAAGCTCCGCGATGAGGTGCTGCTGCCCTT 3043
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
Db 3044 GATGCCAGCGCGGCGCAAAATGCAATGGCTCAGGCAAAAGAAATCTCTAAAGCAACAAAT 3103
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 3104 GTTACCGGATGCTTTAAGAGAGCTTGACGGTACGGCAAAAGCGGTATCGTCTTGGC 3163
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
```

```
Db 3164 ATTCCGACAGAGCTTCACATCCCGTTAAACTAGTCGGTTTAAAGAAACAAAGTTCATGAC 3223
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 3224 CTTACAGAAATTGATCCAGAA-----TCCATGTGTACGGACTC-----TTTTCA 3268
QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 3269 GATTAGTGGAAAAAGCCGAC-----GATTAGAAAAAGGCCCAACATC 3313
RESULT 12
US-08-981-527A-7
; Sequence 7, Application US/08981527A
; Patent No. 6410262
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Kerkman, Richard
; APPLICANT: Broekhuizen, Cornelis P.
; TITLE OF INVENTION: No. 6410262el Secretion Factors for
; TITLE OF INVENTION: Gram-Positive Microorganisms Genes Encoding Them and Methods
; TITLE OF INVENTION: of Using It
; FILE REFERENCE: GCX322-US
; CURRENT APPLICATION NUMBER: US/08/981,527A
; CURRENT FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: PCT/NL96/00278
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1015
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)...(1008)
; OTHER INFORMATION: protein secretion chaperrone
US-08-981-527A-7

Alignment Scores:
Pred. No.: 1,22e-39 Length: 1015
Score: 430.00 Matches: 114
Percent Similarity: 53.29% Conservative: 64
Best Local Similarity: 34.13% Mismatches: 118
Query Match: 18.91% Indels: 38
DB: 4 Gaps: 10

US-09-943-108a-2 (1-455) x US-08-981-527A-7 (1-1015)

QY 4 GluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeu 23
Db 64 GATTCGATCTGAAAGTTTAAAGTATGCGCTTGAAGAAACAAAGAACTCCCTTTCAAAAC 123
QY 24 ThrGluAlaAspIleLysIleMetArgGluValArgLeuAlaLeuPheGlu----- 41
Db 124 AAAGTGAATGATCTGTGTATCCCGTTACCGTAAAGTGGATGAGGATTCCTCGAAGAGCTT 183
QY 42 -----AlaAspValAsnPhelLysValLysGluPheIle----- 53
Db 184 GAAGAGGTTCTTATCAGCGGATGTCGGTTTAAACCGCTTATGAAGTAAATATAGATGAG 243
QY 54 -----LysThrValSerGluArgAlaLeuGlySerAspValMetGlnSerLeuThrPro 71
Db 244 CTGAAAAAAGAGTCAAAACGAGA-----AATATTCAAGAT 279
QY 72 GlyGlnGlnValIleLysIleValGlnAspGluLeuThrLysLeu-----MetGlyGly 89
Db 280 CCAAGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 339
QY 90 GluAsnThrSer-----IleAsnMetSerAsnLysProThrValValMetMetValGly 108
Db 340 GAGCAAAATTCAGAACTGAACATCCAGGATGGCGGTTTAAACGTAATCTCTGTGTAGGT 399
```



```

QY 109 LeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArgLysLys 128
Db 400 GTAAACGGCGTCGGGAAACAAACACGATCGGAAGCTGCTCATAAATG---AAACAA 456

QY 129 TyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArgProAlaAlaIleAsnGln 148
Db 457 GAAGGAATATCTGTACTTGCCTGCGGAGACACTTTAGAGCGGGAGCCATGAAACAG 516

QY 149 LeuGlnThrValGlyLysGlnLeuAspIleProValTyrSerGluGlyAspGlnValLys 168
Db 517 CTGGAAGTATGGGAGACGTCACAGGAGTGCCTGTCTATTACGACAGCGGACGGAACGAT 576

QY 169 ProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAspPheVal 188
Db 577 CCGCGCGCTGTCTATCAGACGCTGCTTCATGCTGCGAAGCAAGAAATGCCGATGTATTA 636

QY 189 IleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLysGlu 208
Db 637 ATTGTGATACGGCAGCGCTCTCCAAACAAAGTAATCTCATGAAGAGCTTGAAAAA 696

QY 209 ValLysGluIleAlaLys-----ProAsnGluIleMetLeuValVal 222
Db 697 GTAACACGCTTATCGAAGAGAGATTCCTGAAGCTCCGATGAGGTGCTGTGCCCTT 756

QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
Db 757 GATGCCACACCGCGGCAAAATGCAATGGCTCAGGCAAAAGAAATCTCTAAGCAACAAAT 816

QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 817 GTTACCGCATGCTTTACGAGACTTGACGGTACGGCAAAAGCGGTATGCTTCCTTGGC 876

QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
Db 877 ATTGCAACGAGCTTCACATCCCGTTAAACTAGTCGGTTTAGAGAGAAAAGTTGATGC 936

QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 937 CTCGGAATTTGATCCAGAA-----TCTATGTGTACGACTC-----TTTTC 981

QY 303 SerLeuIleGluLysAlaGlnAspValAspGlnGluLys 316
Db 982 GATTAGTGGAAAGCCGAC-----GATTAGAAAAA 1014

RESULT 13
US-09-134-001C-1980
; Sequence 1980, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lyon Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1980
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1980

Alignment Scores:
Pred. No.: 4,2e-39 Length: 1248
Score: 426.50 Matches: 103
Percent Similarity: 54.57% Conservative: 76
Best Local Similarity: 31.40% Mismatches: 122
Query Match: 18.76% Indels: 27
Db: 4 Gaps: 8

```

```

US-09-943-108a-2 (1-455) x US-09-134-001C-1980 (1-1248)

QY 8 GluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeuThrGluAlaAsp 27
Db 298 CAAACCTCCCAAGAACAGCTAAATTAATTAATTCCTCGATATAGAAAAGTTGACGAGAT 357

QY 28 IleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAlaAspValAsnPheLys 47
Db 358 TTCTTCGACCTCTGGAAGAATG-----CTTAATTACTGCGGACGTGTGTTTAAT 408

QY 48 ValValLysGluPheIleLysThrValSerGluArgAlaLeuGlySerAspValMetGln 67
Db 409 ACCGTTATGAATTAACCTGATGAGCTACGACAGCAACAGCAACAGCTTAATATACAGAA 468

QY 68 SerLeuThrProGlyGlnGlnValIle---LysIleVal-----GlnAsp 81
Db 469 ACAGAGACTTAAGAGAGTATTAGTTGAGAGATGTAGAAATCTATCATCAAGAGSAC 528

QY 82 GluLeuThrLysLeuMetGlyGlyLysThrSerIleAsnMetSerAsnLysProPro 101
Db 529 GATTATCTCGAAGCAATGAATATTGAAGATGGACGTTAAAT-----570

QY 102 ThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeu 121
Db 571 ---GTCATACTGATGTTGTTGTTGATGTCGCAACAAACAAACAATTTGTTAAATTA 627

QY 122 AlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyr 141
Db 628 GCTTATCGTTTCAACAGAA---GGTAAAAAGTAATGTTAGCCGCTGGTGATCACTTC 684

QY 142 ArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyr 161
Db 685 AGAGCTGGAGCAATTCACAAATTAACGCTCGGGAGAACGTTGTTGTTGAGTTG 744

QY 162 SerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLys 181
Db 745 AGTCAAAACGAAAGTTCTGACCCCTGCACAGCTAGTAGTATGATGCGATTAAATCGCGCAA 804

QY 182 GluGluHisLeuPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAla 201
Db 805 AATAAGCGGTAGATATTTTATTTGTGATCTGCAGGACGCTTGCAAAATAAATCTAAT 864

QY 202 LeuMetAsnGluLysGluValLysGluIleAlaLys-----Pro 215
Db 865 TTAATGCAAGAGCTTAGATAAAATGAACGTTGTGATTAAATCGTGCAATACCTGATGCCCC 924

QY 216 AsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlu 235
Db 925 CATGAAGCTTTTATGCTTGGATGCAACACTGGTCAAAATGCACCTTTCACAGCACGT 984

QY 236 SerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThr 255
Db 985 TCATTTAAGGAAGTTACAAATGCTCAGGTATAGTTTAACTAATATAGACGGTACTGCT 1044

QY 256 ArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGly 275
Db 1045 AAAAGGGGTATTGTTATAGCAATTCGAAATGAGTTACACATTCAGTTAAATATATGTTGT 1104

QY 276 MetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIle 295
Db 1105 TTAGCGCAAAAAATGGATGACTTACACCGTTTAAATCTCTGAA-----AGCTATGTA 1155

QY 296 LeuGlyMet---GlyAspValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGln 314
Db 1156 TATGGATTTATTTGCTGATATGATAGAACAAATGAGATATATCTCTGAAGAAATCTCTAGA 1215

QY 315 GluLysAlaLysAspLeuGluLys 322
Db 1216 AATTATCGGTTGTTGATCTGGAAGAA 1239

RESULT 14
US-08-858-207A-180

```

```

; Sequence 180, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19408-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmel, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 180:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 836 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-858-207A-180

```

```

Alignment Scores:
Pred. No.: 3,34e-39 Length: 836
Score: 425.00 Matches: 83
Percent Similarity: 76.35% Conservative: 30
Best Local Similarity: 56.08% Mismatches: 35
Query Match: 18.69% Indels: 0
DB: 4 Gaps: 0

```

```
US-09-943-108A-2 (1-455) x US-08-858-207A-180 (1-836)
```

```

QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 393 ATGGCATTGAACTTTAAACAGACGCTTTCAGAACGCTTTAAATAATCTACGTAATAAAA 452
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 453 GGAATAAATCTCGAATCTGATGTGCCAAGAGGCAACCAAGAAATTCGCTTGCCCTGTC 512
QY 41 GluAlaAspValAsnPheLysValValLysGluPheLleLysThrValSerGluArgala 60
Db 513 GAGGCGGACGTTGCTGCTGTGTAAGGACTTTATCAAGAAAGTTCGTGAGCGGCA 572
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 573 GTCGGGCTAGGGTCATGTATACACTTAATCTGGCGAACAGATATTATAAATCTGTGAT 632
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100

```

```

Db 633 GAGGAACACAGCCGTTTAGTGTCTGATACGGCAGAAATATCAAGTCACCTAAGATT 692
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 693 CCAACCATCATCATGATGGTTGGTTTACAAAGGGCTGGTAAACAACCTTTGCTGTATAA 752
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 753 TTGGCCACAACACCAAGAAAGAAATATGCTGCTCTTTGATGATTGGCGGGATATT 812
QY 141 TyrArgProAlaAlaIleAsnGln 148
Db 813 TATGTCACCATGCCATTCGACAG 836
RESULT 15
US-09-007-476-1
; Sequence 1, Application US/09007476
; Patent No. 6159949
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6159949el Ftsy
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/007,476
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GMI0079
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1251 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-007-476-1

```

```

Alignment Scores:
Pred. No.: 1.78e-38 Length: 1251
Score: 421.00 Matches: 101
Percent Similarity: 55.59% Conservative: 78
Best Local Similarity: 31.37% Mismatches: 121
Query Match: 18.51% Indels: 22
DB: 3 Gaps: 7

```

```
US-09-943-108A-2 (1-455) x US-09-007-476-1 (1-1251)
```

```

QY 8 GluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeuThrGluAlaasp 27
Db 289 CAAATTTCCAGACAACATTAATAATTTGATAGCGGATATCTGTAAGTAGATGAAGAC 348
QY 28 IleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAlaaspValAsnPhelys 47

```

```
Db 349 TTTTTCAGACCTTAGAAGAAATG-----TTAATCAGTCGAGAGTCGGTTTTTAAT 399
QY 48 ValValLysGluPheLeIleYThrValSerGluArgAlaLeuGlySerAspValMetGln 67
Db 400 ACAGTG-----ATGACGTTAACTGAAGAATTACGTATGGAAGCACACACGCGT 447
QY 68 SerLeuThrProGlyGlnGlnValIleYsIleValGlnAspGluLeuThrLysLeuMet 87
Db 448 AATATTCAAGATACTAGAGATTGCGTAAGTCATTGTTGAAAAAATCGTAGAGATTAC 507
QY 88 GlyGlyGlu-----AsnThrSerIleAsnMetSerAsnLysProProThrValVal 104
Db 508 CATCAAGAAGATGATAATTCAGAGCTATGACATAGAACAGATGCGTGTAAATGTCATT 567
QY 105 MethMetValGlyLeuGlnGlyAlaGlyThrThrAlaGlyLysLeuAlaLeuLeu 124
Db 568 TTAATGTCGCGTGTGAATGCTGGTAAACACACAACTTGGAAATTAGCT---TAC 624
QY 125 MetArgLysLysTyraSnLysLysProMetLeuValAlaAlaAspIleTyArgProAla 144
Db 625 CCATATAAATGAAGGTAAAAAGTAATGTAGCTCGGGCGATACHTTTAGAGCGGGT 684
QY 145 AlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTySerGluGly 164
Db 685 GCTATTGATCAATTGAAGCTTTGGGGCGAAGCTGTGGTGTAGATGTAATTAGCCAAAGT 744
QY 165 AspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHis 184
Db 745 GAAGGTCTGATCCACTGCTGCTATGATGATGCGGATTAATGCCGCTAAAAACAAGGT 804
QY 185 LeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsn 204
Db 805 GTTGATATTTTAACTGTGATACCGCTGGAGCTTTACAAAATAAAACAATCTAATGCCAA 864
QY 205 GluLeuLysGluValLysGluIleAlaLys-----ProAsnGluIle 218
Db 865 GAATTAGAAAAAGTTAAGCGTGTAAATTAATCGACAGTCGCCAGTCGCCCTCATGAAGCA 924
QY 219 MetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAsp 238
Db 925 TTACTATGTTAGATGCTACACTGCTCAGATGCGTGTGTCACAGCTAGAACTTTAAA 984
QY 239 AspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGly 258
Db 985 GAAGTAACAAATGTTACAGTATTGTTAATACGAATTAGATGGTACAGCCAAAGGTGGT 1044
QY 259 AlaAlaLeuSerIleArgSerValThrGlnLysProIleIleYsPheValGlyMetSerGlu 278
Db 1045 ATCGTATTAGCCATTGCTGAATGAATGCACATTCAGTTAAATATGATAGGTTAGGTGAG 1104
QY 279 LysLeuAspGlyLeuGlnLeuPheHisProGluArg-----MetAlaSerArg 294
Db 1105 CAATTAGATGACTTACACCACTTAACCTGAAGTTATGCTACGGCTTATTCGCTGAT 1164
QY 295 IleLeuGlyMetGlyAspValLeuSerLeuIleGlnLysAlaGlnGlnAspValAspGln 314
Db 1165 ATGATTGAACAAAATGACGGAATAACACACAGTGAATAAT---GATCAAAATTGTAACAGAA 1221
QY 315 GluLys 316
Db 1222 GAAGAG 1227
```

Search completed: February 25, 2003, 03:14:46  
Job time : 109 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 02:59:07 ; Search time 89 Seconds  
(without alignments)  
2871.303 Million cell updates/sec

Title: US-09-943-108A-2  
Perfect score: 2274  
Sequence: 1 MAFEGSERLQATMQMRGK.....GKGGKRNQNMKGMNLPF 455

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 442118 seqs, 280819700 residues  
Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlpl  
-Q=/cgn2\_1/USPTO.spool/US09943108/runat\_21022003\_151633\_1953/app\_query.fasta\_1.647  
-DB=PublishedApplications.NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09943108.ecgn\_1\_1\_80\_@runat\_21022003\_151633\_1953  
-NCPU=6 -ICPU=3 -NO\_XLPUX -NO\_MMAB -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications.NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
1	2261	99.4	1368	10	US-09-815-242-8968	Sequence 8968, Ap
2	2235	98.3	1365	10	US-09-815-242-4365	Sequence 4365, Ap
3	2235	98.3	1368	10	US-09-815-242-8030	Sequence 8030, Ap
4	1418.5	62.4	1416	10	US-09-815-242-3890	Sequence 3890, Ap

Sequence 6764, Ap  
Sequence 37, Appli  
Sequence 1847, Ap  
Sequence 9156, Ap  
Sequence 9427, Ap  
Sequence 1, Appli  
Sequence 2255, Ap  
Sequence 6166, Ap  
Sequence 1947, Ap  
Sequence 7850, Ap  
Sequence 6868, Ap  
Sequence 9991, Ap  
Sequence 1, Appli  
Sequence 1766, Ap  
Sequence 1985, Ap  
Sequence 2013, Ap  
Sequence 3070, Ap  
Sequence 877, App  
Sequence 2247, Ap  
Sequence 6281, Ap  
Sequence 369, Ap  
Sequence 8028, Ap  
Sequence 4244, Ap  
Sequence 1852, Ap  
Sequence 279, App  
Sequence 6807, Ap  
Sequence 7684, Ap  
Sequence 3937, Ap  
Sequence 9165, Ap  
Sequence 1, Appli  
Sequence 9491, Ap  
Sequence 6997, Ap  
Sequence 3, Appli  
Sequence 2259, Ap  
Sequence 2506, Ap  
Sequence 1094, Ap  
Sequence 1094, Ap  
Sequence 1094, Ap  
Sequence 38, Appli  
Sequence 13494, A  
Sequence 14727, A

ALIGNMENTS

RESULT 1  
US-09-815-242-8968  
; Sequence 8968, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8968
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8968

Alignment Scores:
Pred. No.: 5,28e-245 Length: 1368
Score: 2261.00 Matches: 452
Percent Similarity: 99.78% Conservative: 2
Best Local Similarity: 99.34% Mismatches: 1
Query Match: 99.43% Indels: 0
DB: 10 Gaps: 0

US-09-943-108A-2 (1-455) x US-09-815-242-8968 (1-1368)

QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGSCAATTGAAGGGTTATCAGACGGCTTGCAAGCGACGATGCAAAAAATGCGTGAAG 60

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAACCTTACTGAACCTGATTAAGATAATGATGCGTGAAGATGAGATTACGCTTACTT 120

QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAAACTTTAAAGTGTAAAGCAATTTATTAACACAGTATCAGACGGCGCA 180

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 TTAGTTCCTGATTAATGCAATCATTAACACAGGCGCAACAGTTATTAATAATAGTTCAA 240

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGATTAACGCGATGTTGAGGGTGGAGAAATACGTCGATTAATATGTCAAATAACCA 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTGTATGATGTTGTTTACAGGTGCTGGTAAACACAACTGCAGGTAAA 360

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACAACTATGTTAGTTCGACGATATT 420

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATGTCACAGCAGCATTAATCAATTAACACAGTACGAGCAAAATGATTCCTGTA 480

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAATTAATGCAATTAACATGCT 540

QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACATTTTATGCTTTGTAATCAATGATACAGCAGGTCGATTACACATCGATGAA 600

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTGATGACGATTAATAAGAGATTAACACATTCCTTAACCAACCAAAATTAATGTTA 660

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATTCATGAGGGTCAAGATGCTGCTCAATGTTGCAAGATCTTTGACGATCAA 720
```

```
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTTACCTTAACATAATAGATGTCATACACAGTGGTGGTCACT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTATCTATTCTGTTGCGGTGACACAAACCAATTAATTTGTTGGTATGAGTGAAGATTGA 840

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGGTTTACAGCTATTCCATCCTCGACGATGSCATCAGTATTATTAGGTATGGGTGAT 900

QY 301 ValLeuSerLeuIleGlnLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 GTGTTAAGTTTAAATGAAAAAGCGCACACAGATGTGGATCAAGAAAAAGCAAAAGATTGA 960

QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluLeuAspGln 340
Db 961 GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTATAGAACAACTTGTATCAG 1020

QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATCTAGGACCACTGGATGATATTATGAAATGATTCAGGTATGAATAAATG 1080

QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
Db 1081 AAAGGGCTAGATTAAGCTTAATATGATGTAAGCAAAATTCATATTAAGCGATTATC 1140

QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACGGCGCTGAAAGAAACAAATCCACACACATTTGAATGTATCAGTAAAG 1200

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTAAAGGCTGCTGCTGCTTTCATTCATCAAGAAAGTCAATCGTTTGCATGA 1260

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysGlyLysLys 440
Db 1261 AACGATATGAAGAAATGATGAACAGTTCAGTGGTGGGTAAGGTAAGGTAAGGTAAG 1320

QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMetAsnLeuProPhe 455
Db 1321 CCAATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365

RESULT 2
US-09-815-242-4365
; Sequence 4365, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
```

; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4365  
; LENGTH: 1365  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-4365

## Alignment Scores:

Pred. No.: 4,42e-242 Length: 1365  
Score: 2235.00 Matches: 447  
Percent Similarity: 98.90% Conservative: 3  
Best Local Similarity: 98.24% Mismatches: 5  
Query Match: 98.28% Indels: 0  
DB: 10 Gaps: 0

US-09-943-108A-2 (1-455) x US-09-815-242-4365 (1-1365)

QY 1 MetAlapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
DB 1 ATGGCAATTTGAAGGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATGCGTGTAAAG 60  
QY 21 GlyLysLeuThrGluAlaAspLeuLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
DB 61 GGTAACCTACTGAGCGATGATATAAGATATATGCTGGAAGTAAGATTAGCGTTACTT 120  
QY 41 GluAlaaspValasnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
DB 121 GAGCGTCAGCTAAACTTTAAAGTGGTAAAGAAATTTATTAACAACAGTATCAGAACGCGCA 180  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
DB 181 TTAGGTTCGATGTAATGCAATCATTAACACCGAGGCAACAAGTTATTAAATAGTCAA 240  
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
DB 241 GATGAATTAACGAGCTGATGGGTGGAGAAATATATATGATTAATGTCAAATAACCA 300  
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
DB 301 CCTACTGTGTTATGATGGTGGTTTACAAGTCTGCTGTAACAAACAACACTGCAGGTAAA 360  
QY 121 LeuAlaLeuMetArgLysThrAsnLysLysProMetLeuValAlaAlaAspIle 140  
DB 361 TTAGCATTTATGCGTAAAGAAATACACAAACAAACCTATGTTGTCAGCAGATATT 420  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
DB 421 TATCGTCCAGCAGCGATAATCAATTAACAACAGTAGGGAACAATAATGATATCTCTGTA 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
DB 481 TACAGTGAAGGAGTCAAGTAAGAACCCACAACAATTTGTAACTAATGCAATTAACATGCT 540  
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200  
DB 541 AAGAAGAATTTAGACTTTGTAATCATTTACAGCAGGTCGGAATACATCTCGATGAA 600  
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
DB 601 GCATTTGATGAATGAATTAAGAAGTAAGACATTTGCTAAACCAACAAAGAAATTATGTTA 660  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
DB 661 GTTCTGATTCATGACGGGTCAAGATCGTCAATGTTGCAAGATCTTTTGACGATCAA 720  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260  
DB 721 CTTGATGTCAGGTTTACCTTAACATAAATATAGATGGTGAACACGCTGGTGGTCAGCT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
DB 781 TTATCATTTTCGTTTCGGTGACACAAAACCAATTAATTTGTTGTTATGAGTGAAGATT 840  
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
DB 841 GATGGTTTAGAGCTATTCCATCTCGAACGATGATGGCATCACGATTTTCAGCGCTGGGTGAT 900  
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
DB 901 GTCTCAAGGTTAATTGAAAAAGCCACACAGATGTGGATCAAGAAAAAGCAAGATT 960  
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
DB 961 GAGAAAAAGATCGCGAGTCATCAATTACTTTAGATGATTTTGTAGAACACCTTGATCAG 1020  
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
DB 1021 GTGAAAAATTTAGGACCACCTGGATGATATTATTAAAAATGATTCAGGTATCAATAAATG 1080  
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
DB 1081 AAAGGCTAGATAAGCTTAATAGTGAAGAAACAAATTTGATCATATTAAAGCGATTATC 1140  
QY 381 GluSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
DB 1141 CAGTCAATGACCGCGCTGGAAGAAACAATCCAGACACATTTGATGTATCACGCTAAAG 1200  
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
DB 1201 CGTATGCTAAAGGTTCTGGTCTCATTAACAAGAAGTCAATCGTTTGATCAACAATTT 1260  
QY 421 AspAsnMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440  
DB 1261 AACATATGAAGAATAATGATGAACAATTCCTGTCGGCGGTAAGGTAAAGGTAAAGTAA 1320  
QY 441 ArgAsnGlnMetGlnAsnMetLysLysGlyMetAsnLeuProPhe 455  
DB 1321 CGCAATCAAAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365

## RESULT 3

US-09-815-242-8030  
; Sequence 8030, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8030
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8030
```

```
Alignment Scores:
Pred. No.: 4,44e-242 Length: 1368
Score: 2335.00 Matches: 447
Percent Similarity: 98.90% Conservative: 3
Best Local Similarity: 98.24% Mismatches: 5
Query Match: 98.28% Indels: 0
DB: 10 Gaps: 0
```

US-09-943-108A-2 (1-455) x US-09-815-242-8030 (1-1368)

```
QY 1 MetAlaPheGluGlyLeuSerGluGluGluGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCATTTTGAAGCTTATCAGAACGGCTGCAAGCGACGATGCAAAAAATCGTGTAAG 60
QY 21 GlyLysLeuThrGluAlaAspPheLysLysMetMetArgGlnValArgLeuAlaLeuPhe 40
Db 61 GGTAACTTACTGAGCTGATATAGATATATGATGCTGAAGTAGATTAGCGTTACTT 120
QY 41 GluAlaAspValAsnPhelysValLysGluPheLysThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAAACTTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGGGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysValGln 80
Db 181 TTAGGTTCGATGTAATGCAATCATTAACACGAGGCAAGATTTATTAATTAAGTTCAA 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGATTAACGAAGTTGATGGTGGAGAAAATATATCGATTATATGTCAAATAAACCA 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTGTTATGATGGTGTGTTTACAAAGTGTGCTGTAACAAACAACTCCAGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTATGATGCGGTAAATAATACACAAAACAACTATGTTGTCACAGATATT 420
QY 141 TyrArgProAlaAlaLalleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGAAACAAATTTGATTTCTCTGA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAGGAGATCAGTAAAGCAACAAATGTAATGTAATGCAATTAACAACTGCT 540
QY 181 LysGluGlnHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisLysAspGlu 200
Db 541 AAAGAGAACAATTTAGACTTTGTAATCATTTGATACAGCAGCGTGGATTACACATCATGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLysLysProAsnGluLeuMetLeu 220
Db 601 GCATTGATGATGATTAATAAGAAAGTAAAGACATGCTAAACCAACAAATTTATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATTCATGACGGGTCAAGATGCTGTCATGTTGCAGAACTTTTGCATCAAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTTACCTTAACATAAATTAGATGCTGATACAGTGTGTCAGCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280
```

```
Db 781 TTATCTATTGTCGTCGTCGACACAAAAACCAATTAATTTGTTGGTATGAGTAAAAAGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGGTTTGAAGCTATTCATCCTGAACGATGATGCATCAGCTATTTCAGGCTGGGTGAT 900
QY 301 ValLeuSerLeuLleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTCAAGGTTTATTTGAAAAACGCAACAGATGTGGATCAAGAAAAAGCAAGATTTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
Db 961 CAGAAAAAGATGCCGAGTCATCATTTTACITTAGAAGATTTTACACACACTTGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATTTAGCACCACTGGATGATATTATTAATAATGATTCAGGATGATGAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisLysAlaLlelle 380
Db 1081 AAAGGTCTAGTAAGCTTAATATGATGAGTGAAGCAAAATGATCATATTAAAGCGATATC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACGCGCGCTGAAAGAAACAATCCAGACACATTCGATGATCAGTAAAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTCTTAAAGGTTCTGGTCTCATTAACAAGAGTCAATCGTTGATGAACAAATTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysGlyLysGlyLys 440
Db 1261 AACATATGAGAAATGATGAACAATTCACITGTCGCGGTAAAGGTAAAGGTAAAGGTAAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CCAATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365
```

#### RESULT 4

```
US-09-815-242-3890
; Sequence 3890, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 3890
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3890

Alignment Scores:
  Pred. No.:      3,17e-150      Length:      1416
  Score:          1418.50        Matches:     273
  Percent Similarity: 77.54%     Conservative:  86
  Best Local Similarity: 58.96%   Mismatches:   89
  Query Match:     62.38%       Indels:       15
  DB:              10          Gaps:           2

US-09-943-108a-2 (1-455) x US-09-815-242-3890 (1-1416)

QY 1 MetAlaPheGluLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 ATGGCTTTTGAGAGTTTACAAAACGGCTACAAACAGGCAATGAGTAAATCCGTCGTAAG 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 21 GlyLysLeuThrCluAlaAspLysLysMetMetArgGluValArgLeuAlaLeuPhe 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 GGAAGAGTTCCGAGCGCGAGTAAAGAAATGATCGAGAAATCCGTTGGCTTATTA 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 41 GluAlaAspValAsnPhelLysValLysGluPheLysThrValSerGluArgAla 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 GAAGCGAGCTTAATTTACAAAGTGTCAAGATTTTCCAAAACGGCTCAGAGAACGGGCA 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLeValGln 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 GTAGGAGTCGAAGTATTAGAAGCTTATCACCAGCCCAACAAATTTGTAATAATTTGTTGAT 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 81 AspGluLeuThrLysLeuMetGlyGlyLysThrSerLysAsnMetSerAsnLysPro 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 GAGAGATTACGAAAGCTTAGGTTCAGAAACGGTTCAACTGAATAAATTCCTCAAAAATC 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 301 CCGACAGTGATATGATGACAGGGTTACAGGGGCTGGTAAACAACTTTTACTGGTAA 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 361 TTACAAAACACTTAATGAAAACGAAACGGCTCGTCCGCTTTTAAATCGCTGGTGACGTT 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 141 TyrArgProAlaAlaAlaLeuGlnThrValGlyLysGlnLysLysLysLysLysLys 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 421 TATCGTCCAGCAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLysValThrAsnAlaLeuLysHisAla 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 481 TTTCATATGGACAGACATGCTTAATCCAGTGGAAATTTGTCGTCAGGGTTAGCATAGCA 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 181 LysGluGluHisLeuAspPheValLysLysThrAlaGlyArgLeuHisLysLysLysLys 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 541 AAAGAAAAGAAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLysLysLysLysLysLysLysLys 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 601 GCITTAATGACGAATTAAGAAATTAAGAGTTTGGCTTAATCCCAATGAAATTCGTGTA 660
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 221 ValValAspSerMetThrGlyGlnAlaValAsnValAlaGluSerPheAspAspGln 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 661 GTTGTGATGCGATGACGGGCGAGATGCTCAAGTTGTCAGATAGTTTATGAACAG 720
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 721 CTTGCAATTAATCGGGTGTATTATACCAAAATGACGGCGATACGCTGGGGGGGCGTGC 780
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 781 CTGTCAATTCGGGCGATGACGGGCGCTCCGATTAATTTGTCGGTTCGTGTAATAATTA 840
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgLysLeuGlyMetGlyAsp 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```



; SEQ ID NO 6764

; LENGTH: 1434

; TYPE: DNA

; ORGANISM: Enterococcus faecalis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)....(1434)

US-09-815-242-6764

## Alignment Scores:

Pred. No.: 3,23e-150 Length: 1434  
 Score: 1418.50 Matches: 273  
 Percent Similarity: 77.54% Conservative: 86  
 Best Local Similarity: 58.96% Mismatches: 89  
 Query Match: 62.38% Indels: 15  
 DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-815-242-6764 (1-1434)

QY 1 MetAlaPheGluGlyLeuSerGluArgGluGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCTTTGAGAGTTTAAACAACCGCTACCAACAGGCATGAGTAAATCCGTCGTAAAG 60  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 61 GGAAAGTTTCGGAAGCGGACGTAAAGAAATGATGCGAGAAATCCGTTGGCTTTATTA 120  
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 121 GAAGCCGAGGTTTAAATTCAGAGGGTCAAGATTTTCACAAACCGCFCAGAGAGCGGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 Db 181 GTAGGAGTCGAGATTTAGAAAGCTTATCACCGACCCCAACAATGTGAAAATTTGTGAT 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 241 GAAGAAATTAACGAAACGTTAGGTTCAGAAACGGTTGAACGTAATAATCCCAAAATC 300  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCACAGTGATTTATGATACAGGGTTACAGGGGCTGGTAAACAACTTTTACTGGTAAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrosLysLysProMetLeuValAlaAlaAspIle 140  
 Db 361 TTAGCAAAACACTTAATCAAACTGAAACGCTCGCTCTTTTAAATCGTGGTGACGTT 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 421 TATCGTCCAGCAGCGATTGATCAGTTGAAGGTTTGTAGTCAACAATTAGAAGTTCCCGTT 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTTGATATGGACAGAGCTTAATCCAGTGGAATGTTCGTCAAGGTTTAGCATTAGCA 540  
 QY 191 LysGluGluHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
 Db 541 AAAGAAAGAAATGATTGTCTTAATGATAGCGCGCGCTTTTACACATTGACGAA 600  
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 GCTTTAATGAGCAATTAACAAATTAAGAGTTGGCTATCCCAATGAATTCGTGTA 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
 Db 661 GTTGTGATCGATGACGGGCGAGATGCTGTCACGTTGCAGATAGTATTAATGAACAG 720  
 QY 241 LeuAspValThrClyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 721 CTGTGAATTAAGGGTGTGTTATTACCAAAATGACGCGCATCTCGTGGGGGGGCTGG 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280

Db 781 CTGTCAATTCGGCAGTAAACAGGGCGCTCCGATTAAATTTGTCGGTTCTGGTGAATAATTA 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 ACCGATTTAGAAATTTTCATCCGATCGTATGCGAGTCGTATAGGTATGGGGAC 900  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 Db 901 AFGTIGAGCTTAATGAAAAACGCCACAGATACGATGAGAAAAAGCAGAAAGACTT 960  
 QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
 Db 961 GGTCAAAAAATGAAGAAACAGTTTTCGACTTAAACGATTTCATTGACCAATTGATCAA 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 1021 GTATGGCATGGGACCGGATTAAGACCTTATTAATAATGATCCCTGGAATGAGTAACATG 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380  
 Db 1081 CCTGTTATGAAATGTCAAAAGTCGATCCAAAGATGTGGCAGGAAACGGCGATGGTC 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 CTATCAATGACCCCTGCAGACGTCGAAATCTGATCTATTAAATCCATGCGCCGTCGC 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 AGAATTCAGCTGGTTCAGGAATAGTGTGGTGAAGTCAATCGTATGATTAACAAATTT 1260  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThr----- 431  
 Db 1261 AAAGAAATCAAAAAATGATGCAACAAATGTCCAAAGGGGATATGAACATTCCTGGTATG 1320  
 QY 432 -----GlyGlyGlyLysGlyLysGlyLysGlyLysArgAsnGlnMetGlnAsn 446  
 Db 1321 GATCAATGCTAGTGGCGCGCTTAAGCAAGTATAGTTAAATG---GCCATGATCGT 1377  
 QY 447 MetLeuLys 449  
 Db 1378 ATGATGAAG 1386

RESULT 6  
 US-09-070-927A-37  
 ; Sequence 37, Application US/09070927A  
 ; Patent No. US20020120116A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles A. Kunsch  
 ; Steven Barash  
 ; Patrick J. Dillon  
 ; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
 ; NUMBER OF SEQUENCES: 982  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/070,927A  
 ; FILING DATE: 04-May-2000  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION NUMBER:  
 ; APPLICATION NUMBER: 60/046,655  
 ; FILING DATE: 1997-05-16  
 ; APPLICATION NUMBER: 60/044,031  
 ; FILING DATE: 1997-05-06

; APPLICATION NUMBER: 60/066,009  
 ; FILING DATE: 1997-11-14  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kenley K. Hoover  
 ; REGISTRATION NUMBER: 40,302  
 ; REFERENCE/DOCKET NUMBER: PB369  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 37:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6729 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 37:  
 US-09-070-927A-37

Alignment Scores:  
 Pred. No.: 3,83e-149 Length: 6729  
 Score: 1417.50 Matches: 273  
 Percent Similarity: 77.54% Conservative: 86  
 Best Local Similarity: 58.96% Mismatches: 89  
 Query Match: 62.34% Indels: 15  
 DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-070-927A-37 (1-6729)  
 QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 3722 ATGGCTTTTGAGAGTTTAACAAACCGCTACAGCAGCAATGAGTAAATCCGCTGTAAG 3781  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 3782 GGAAGAAGTTCCGAAGCGCAGTAAAGAAATGATCCGAGAAATCCGTTGGCTTATTA 3841  
 QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 3842 GRAGCGGAGTTTATTCACAGTGGTCAAGATTTCAACAAACGCTGTCAGAGACGGGCA 3901  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 Db 3902 GTAGGAGTCGAAGTATTAGAAGCTTATCACCAGCCCAACAAATGTGTAATAATTTGTGAT 3961  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 3962 GAGAAATTAAACGAAACCTTAGGTTCAGAAACGGTTGAACATGAATAAATCTCCAAATATC 4021  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 4022 CCGACAGTCATTATGATCAGCAGGGTTACAAGGGGCTGCTAAACAACTTTTACTGGTAA 4081  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
 Db 4082 TTAGCAAAACACTTAATGAAACATGAAACCGCTCGCTGCTTTAATCGTGTGACGTT 4141  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 4142 TATCGTCCAGCAGGAGTTGATGAGTGAAGTTTATGATCAACAAATGAGAAGTCCCGTT 4201  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 4202 TTTGATATCGGACAGATGCTATCCAGTGGAAATGTTCTCAAGGTTAGCATTAGCA 4261  
 QY 181 LysGluLysLeuLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200  
 Db 4262 AAAGAAAAGAAATGATTATGCTTAATGATACGGCCGCCGCTTTACACATGACGAA 4321  
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluAlaIleAlaLysProAsnGluIleMetLeu 220  
 Db 4322 GCTTAATGGCAATTAACAAATTAAGAGTTGGCTGATCCCAATGAATTCGTGTA 4381  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240

Db 4382 GTGTGTGATGCGATCAGCGGSCAAGATGCTGTCAACGTTGCAGATACTTTTATGACAG 4441  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 4442 CTTGGAATTAAGTGGGTGTTTATTACAAATGATGCGATACCTCGTGGGGGGGTGG 4501  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 Db 4502 CTGTCAATTCGGGCAGTAACGGGTGCTCCGATTAATTTGCGTTCTGTGAAAAATTA 4561  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 4562 ACCGATTAGAAATTTCCATCCGATCGTATGTCGATCGTATCCTAGTAGTGGGGAC 4621  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 Db 4622 ATGTTGACGCTAATGAAAAAGCGCAACAGATTACGATGAGAAAAAGCAGAAAGACTT 4681  
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
 Db 4682 GCTCAAAATGAAAGAAACAGCTTTTGACTTAAACGATTTTCATTGAGCAATTTGGATCAA 4741  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 4742 GTTATGGCAGTGGGACCGCATGAGACTTATTAATAATGATCCCTGGATGACATAAG 4801  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLeu 380  
 Db 4802 CTGTGATTGAAATGCAAAAGTCGATCCAAAGAGTGTGGCAGCAAAACGGCGATGTC 4861  
 QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 4862 TTATCAATGACCCCTGCAGAACGTAATACTCTGATCTATTAAATCTAGTCGCCCGC 4921  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
 Db 4922 AGAATTGACGCTGTTCTCAGGAATAGTGTGGTGAATCAATCGTATGATTAACAATTT 4981  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThr----- 431  
 Db 4982 AAAGAAATCCAAAAATGATGCAACAAATGCAAAAGGGGATATGAACATTCCTGGTATG 5041  
 QY 432 -----GlyGlyGlyLysGlyLysGlyLysGlyLysGlyLysGlnMetGlnAsn 446  
 Db 5042 GATCAAAATCTAGTGGCGCGGTTAAAGGCAAGCTAGGTAAATG---GCCATGAATCGT 5098  
 QY 447 MetLeuLys 449  
 Db 5099 ATGATGAAG 5107  
 RESULT 7  
 US-09-974-300-1847  
 ; Sequence 1847, Application US/09974300  
 ; Patent No. US20020146721A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berka, Randy M.  
 ; APPLICANT: Clausen, Ib Groth  
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 ; TITLE OF INVENTION: Expression  
 ; FILE REFERENCE: 10085.500-US  
 ; CURRENT APPLICATION NUMBER: US/09/974,300  
 ; PRIOR APPLICATION NUMBER: 09/680,598  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: 60/279,526  
 ; PRIOR FILING DATE: 2001-03-27  
 ; NUMBER OF SEQ ID NOS: 8481  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1847  
 ; LENGTH: 1136  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus licheniformis

US-09-974-300-1847

## Alignment Scores:

Pred. No.: 1,66e-148 Length: 1136  
Score: 1402.00 Matches: 270  
Percent Similarity: 87.30% Conservativity: 60  
Best Local Similarity: 71.43% Mismatches: 48  
Query Match: 61.65% Indels: 0  
DB: 10 Gaps: 0

US-09-943-108a-2 (1-455) x US-09-974-300-1847 (1-1136)

QY 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
DB 1 ATGGCATTTCGAAGGATTAGCAGCGACTGACGCAACAGCATCTCAAAAATCCGGGAAA 60  
QY 21 GlyLysLeuThrGluAlaAspPheLysLeuMetArgGluValArgLeuAlaLeuPhe 40  
DB 61 GGAAGAGTAAAGGATGAGCAAGCTCAAGAGATGATGGCGAAGTCGGTTTACGCTCCTT 120  
QY 41 GluAlaAspValAsnPhelYsValLysGluPheLysThrValSerGluArgAla 60  
DB 121 GAAGCGGAGCTCAATTTAAAGCTGCGAAGATTTTGTGAAAGGTAAGCGACGAGCT 180  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLeuLysLeuValGln 80  
DB 181 GTGGACAAGACGTTATGAAAAGCTGACGCCCGGACAGCAGGTCTATTAAAGTGTAATA 240  
QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerLeuAsnMetSerAsnLysPro 100  
DB 241 GAAGAGCTGAGGAGCTGATGGCGGCGGAGAAAGATTCGCGTGCAGAACGCGCCG 300  
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
DB 301 CGGACGCTCATCATGATGCTGGTCTGCAAGGGCGGTAACACGCTGCGACCGGAAG 360  
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspPhe 140  
DB 361 CTTGCCAACCTGCTCGCAAAAACATTAACCGCAATCCGCTGCTGCGACAGACATC 420  
QY 141 TyrArgProAlaAlaLeuGlnGlnLeuGlnThrValGlyLysGlnLeuAspPheVal 160  
DB 421 TACCGCGCGGCGGCGATCAACACAGCTGGAAACGCTCGCAAGCAGCTGATATCCGGTA 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180  
DB 481 TTTTCACTGGGTGACAGCTGAGTCCGCTGGAGATTCACAGCAGCCATCGAAAGCG 540  
QY 181 LysGluGluHisLeuAspPheValLeuLeuAspThrAlaGlyArgLeuHisLeuAspGlu 200  
DB 541 AAAGAAGATCATCAGCACTACGTTCTCATATACGCGGCGGCGCTTCATATTGATGAA 600  
QY 201 AlaLeuMetAsnGlnLysGlnValLysGlnLeuLeuAlaLysProAsnGluLeuMetLeu 220  
DB 601 GAGCTGATGAGTATGAGCTGAGCAGTAAAGAACCCCTCAGCCGGAAGAGATTTCCTT 660  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAlaAsnValAlaGluSerPheAspGln 240  
DB 661 GTCGTTGACTCATCAGCGGTCAAGACGCGTCAATGTTGCCAAAGCTTTAACGACAC 720  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260  
DB 721 CTCGCTGTGACAGGTGTCATCTTGAAGAGCTTGACGCGCATACTAGAGCGGAGCGGCC 780  
QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280  
DB 781 TTGTCAATCCGCGCGGAGAACACGCGCATCAATTCGCCGGAATGGCGGAAGGCTT 840  
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgileLeuGlyMetGlyAsp 300  
DB 841 GATGCGCTGAGCGGCTTTCATCTCTGACAGGATGCTCCGCGGATATCTCGAATGGCGAT 900  
QY 301 ValLeuSerLeuLeuGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysLeu 320

DB 901 GTGCTCACCTTATCGAGAAAGCCCGAGCCCAAGCTAGACGAAGAAAGCGAGGAATC 960  
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340  
DB 961 GAGCAAAAATGAAATAATGAGCTTTACGCTCGACGACTTTTGGAAACAGCTCGGACAG 1020  
QY 341 ValLysAsnLeuGlyProLeuAspPheLysMetLysMetLysProGlyMetAsnLysMet 360  
DB 1021 GTCGCAATATATGGCGCTTGAAGACCTGATCAATGATGCTCCGCGGCGAGCAAAATG 1080  
QY 361 LysGlyLysAspLysLeuAsnMetSerGluLysGlnLeuAspHisLeuLysAla 378  
DB 1081 AAAGCTTAAAAAACGTGAAGGTTGATGAAAACAGCTCAGCCATATCGAAGCG 1134

## RESULT 8

US-09-815-242-9156  
; Sequence 9156, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9156  
; LENGTH: 1572  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1572)  
US-09-815-242-9156

Alignment Scores:  
Pred. No.: 1,28e-137 Length: 1572  
Score: 1307.00 Matches: 253  
Percent Similarity: 74.73% Conservativity: 90  
Best Local Similarity: 55.12% Mismatches: 102  
Query Match: 57.48% Indels: 14  
DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-815-242-9156 (1-1572)

QY 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
DB 1 ATGGCATTTCGAAGGATTAGCAGAGCTTTGACAGACGCTCTTTAAAAATCTACGTAATAAAA 60  
QY 21 GlyLysLeuThrGluAlaAspPheLysLeuMetArgGluValArgLeuAlaLeuPhe 40

Db 61 GGAARAATCTGATCTGATGTCACAGAGCGCAACCAAGAAATTCGCTTGGCCCTGC 120  
Qy 41 GluAlaAspValAspPheLysValValLysGluPheLysThrValSerGluArgAla 60  
Db 121 GAGCGGAGCTTGCCTTGGTGTAAAGAGCTTTATCAAGAAATTCGTCGAGCGTGA 180  
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLeuValGln 80  
Db 181 GTCCGGCATGAGGTCAATGATACACTTAATCTCGCGCAACAGATATATAAATCGTGTAT 240  
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
Db 241 GAGAACTGACAGCGCTTTAGGTCTGATACCGCAGAAATATCAAGTCACCTAGAT 300  
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
Db 301 CCACCATCATCATGATGTTGGTTTACAAAGGGCTGTAAACAACCTTGTGCTGATAA 360  
Qy 121 LeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
Db 361 TTGGCCAAACAACCTCAGAAAGAAATGCTGTCCTTTGATGGTTGGCGGAGATAT 420  
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
Db 421 TATCGTCCAGCTGCCATTAAGACCTTAAGACCTTGGGACACAGATGATGTGCTGTC 480  
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
Db 481 TTTGCACTTGGCAAGAAAGTACCAGCTGTTGAGATGTACGTCAAGTGTGGAGCAAGCC 540  
Qy 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
Db 541 CAACTAAATCAATACGATATGCTTGATGATGACGCGGTCGTTGTCAGATGTAGAG 600  
Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
Db 601 CTCCTCATGAATGAGTTCGTGATGTAAGAGCATTTGGCTCAACCAATGAATCTTGCT 660  
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
Db 661 GTCGTGTATGCTATGATGGTCAGAGAGCAACCAATGTTGGCGGTGATTAATGCTCAG 720  
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260  
Db 721 TTGGAAGTCACTGGGTGATCCTTACCAGATGATGAGCGCATACGTCGTGGTGTGCT 780  
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
Db 781 CTGCTGTCGTGCATACCTACTGGAACCAATCAAGTTCACGTGTACAGGTGAAAGATT 840  
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
Db 841 ACGCAATGAACCTTCACCCAGACCGCATGCTAGCCGTATCTGCTGATGGGGAT 900  
Qy 301 ValLeuSerIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320  
Db 901 ATGCTCACTTTGATGAGAAAGCTTCAGGAATACGATGACAAACAAAGCCCTTGAATG 960  
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
Db 961 GCTGAGAAGATCGCGAAACACACCTTTGATTTAATGATTCATCGATCAATAGATCAG 1020  
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
Db 1021 GTCCAAATATGGGCCCATGTAAGAGACTTGCTCAGATGATTCAGGTATGCCCAACAT 1080  
Qy 361 LysGlyLeuAspLysLeuMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
Db 1081 CCAGCCCTTCAAAACATGAAGTGGATGAACGCCAGATTCGCTGAACGTCGCTATG 1140  
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400

Db 1141 TCTTCGATGACACCTGAGAGCGTGAAAAACCCACAGATTTCTTAATCCAGCCGTGCCGT 1200  
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
Db 1201 CGTATGCTGCTGGTTCGGAATACATACATCTGCTCAAGTCAATAAATTCATCAAGGACTTT 1260  
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLysLysGlyLys 440  
Db 1261 AACAGCTAAACACACTCATCAGCGGTGTTATCTCTGGG----- 1299  
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454  
Db 1300 -----GATAGTAATAAATGATGAACAAATGGGATTAATCCAAATACCTTCCT 1350  
RESULT 9  
US-09-815-242-9427  
; Sequence 9427, Application US/09815242  
; Patent No. US20030061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9427  
; LENGTH: 1395  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1395)  
US-09-815-242-9427  
Alignment Scores:  
Pred. No.: 1.39e-137 Length: 1395  
Score: 1306.00 Matches: 252  
Percent Similarity: 74.73% Conservative: 91  
Best Local Similarity: 54.90% Mismatches: 102  
Query Match: 57.43% Indels: 14  
DB: 10 Gaps: 2  
US-09-943-108A-2 (1-455) x US-09-815-242-9427 (1-1395)  
Qy 1 MetalAlaPheGluGlyLeuSerGluArgLeuGluAlaThrMetGlnLysMetArgGlyLys 20  
Db 1 ATGCGATTTGAAGATTTTAAAGAACGTTTCAGACACGCTTTTAAAAATCTACGTAAAAA 60  
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

```
Db 61 GGAAAAATCTCTGAATCTGATGTCACAGAGCAACCAAGAAATTCGCTTGGCCCTGCTC 120
QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGGCCAGCGTGGCTGCTGTTGTAAGGACTTTATCAAGAAAGTTCGTGAGCGTGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysValGln 80
Db 181 GTCGGCATGAGTGCATGATCACTTAATCTCGCAACAGATATTAAAAATCGTGCAT 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGlnSerIleAsnMetSerAsnLysPro 100
Db 241 GAGGAATGACGCGTTTATAGTTCTCATACGCGCAAAATATCAAGTAAAGATT 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCACCATCATCATGATGTTGTTTACAGGGCTGTGTAACAGACCTTTGCTGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAspIle 140
Db 361 TTGGCCACAACTCAAGAAAGAAAGAAATCTGCTCTTGTGATGATTCGGCGCGATATT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATGCTCAGCTGCCATTGACAGCTTAAGACCTTGGGCAACAGATGATGTCCTGTC 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTTCGACTTGGACAGAGTACCAGCTGTGAGATTGTAGCTCAAGTTTGGAGCAGCC 540
QY 181 LysGluGlnLysLeuAspPheValIleAspThrAlaGlyArgLeuHisLysPdu 200
Db 541 CAAACTAATACATACAGCATCTCTGATTGATGATCTGCGGCTGCTTGCAGATTGATGAG 600
QY 201 AlaLeuMetAsnGluLeuLysGlnValLysGluLeuAlaLysProAsnGluIleMetLeu 220
Db 601 CTCCTCATGATGAGTCTGCTGATGTAAGACATTTGGCTCAACCAATGAATCTTGCT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTCGTTGATGCTATGATTGGTCAGAACAGCCCAATGTTCCGCGTGAGTTTAAGTCTCAG 720
QY 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 721 TTGGAAGTGAATGGGGTGCATCTTCCCAAGATTGATGGCGATCTGCTGTTGGTGTGCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGCTGTGTCATCATATTACTGGAAACCAATCAAGTTCAGTGTACAGGTGAAAAGATT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACGGACATTGAACCTTCCACCCAGACGCGATGCTAGCCGATCTCTGGTATGCGGGAT 900
QY 301 ValLeuSerLeuIleGlnLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 ATGCTCACTTTGATGAGAAAGCTCTCAGGAATACGATGACAAACAAAGCCCTTGAATG 960
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
Db 961 GCTGAGAGATGCGCGAAACACCTTTGATTTTAAATGATTCATCAATTAATGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGCAAAATATGGGCGGATGGAAGACTTGCTCAAGATGATTCAGGATGATGCCAACAAT 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCACTTCAAAACATGAAGTGGATGAACGCCAGATTGCTCGTAAACGTCCTCATTTG 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACATGAAGAACCTGAAACCCAGATTGTTAAATCCAAAGCCGTCGCGGT 1200
```

```
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTGCTGGTTCGGAATACATCTCGAAGCTCAATAATTTATCAAGACCTTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
Db 1261 AACCAAGGCTAAACAGCTCATGCGAGGCTGTATGCTGGG-----1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAAATGATGAAGCAATGGGATTAATCCAAATAACCTCTCT 1350
RESULT 10
US-09-814-041A-1
: Sequence 1, Application US/09814041A
: Patent No. US20020103104A1
: GENERAL INFORMATION:
: APPLICANT: CHEEVER, CHRISTY
: APPLICANT: FECTEAU, DOUGLAS A.
: APPLICANT: LI, HU
: APPLICANT: PAYNE, DAVID J.
: APPLICANT: STEEL, ANGELA
: APPLICANT: WANG, LEI
: TITLE OF INVENTION: METHODS USING THE SRP POLYNUCLEOTIDES
: AND POLYPEPTIDES COMPOUNDS MODULATING THEIR ACTIVITY
: FILE REFERENCE: GM50069
: CURRENT APPLICATION NUMBER: US/09/814,041A
: CURRENT FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/191,008
: PRIOR FILING DATE: 2000-03-21
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 1569
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
US-09-814-041A-1
```

```
Alignment Scores:
Pred. No.: 1569
Score: 1297.00
Percent Similarity: 74.51%
Best Local Similarity: 54.68%
Query Match: 57.04%
Indels: 14
Gaps: 2
```

US-09-943-108A-2 (1-455) x US-09-814-041A-1 (1-1569)

```
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCAATTGAAAGTTTAAACAGACGCTTTCGAGAACGCTTTTAAATAATCTAGTAAATA 60
QY 21 GlyLysLeuThrGluAlaAspIleLysMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAATAATCTGTAATCTGATGTCCTCAAGAGCAACCAAGAAATTCGCTTGGCTTGCCTC 120
QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGGCGGACGTTGCTGCTGTTTAAAGGACTTATCAAGAAAGTTCGTGAGCGTGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTCGGGCATGAGGTCATGATACACTTAATCTCGCAACAGATTTATTAATCGTTGAT 240
QY 81 AspGluLeuThrLysLysMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAGGAACGTACAGCGGTTTAGGTTCTGTATACGCGCAAGAAATTTATCAAGTCACTAAGATT 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCACCATCATCATGATGTTGTTTACAGGGCTGTGTAACCAAGCCGTTTCTGCTGGTAAA 360
```

QY 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140  
Db 361 TTGGCCAAACAARACTCAAGAAAGAGAAAATGCTGCCTTTGATGATGGCGGATATT 420  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
Db 421 TATCGTCCAGCTGCCATTGACCACTTAAGACCTTGGGACACAGATTGATGTCCTGC 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
Db 481 TTGGCACTTGGACAGAAAGTACCAGCTGTGTGAGATTGTACGTCAAGTTTGGACCAAGCC 540  
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
Db 541 CAACAAATCAATACAGATATGTTGATTGATACCTCGGGTCTGACAGATTGATGAG 600  
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220  
Db 601 CTCCTCATGAATGAGCTTCGTGATGTGAAGATTGCTGCTCAACCAATGAATCTTGCTT 660  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
Db 661 GTCGTTGATGCTATGATGCTGCAGAGACCAAGCAATGTGGCGCTGATTTAATGCTCAG 720  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260  
Db 721 TTGGAGTGCATGGGGTTCATCTTACCAAGATTGATGCGATACCTGCTGCTGCTGCT 780  
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
Db 781 CTGCTGCTGCTGCATCATCTGGAACCAACCAATCAAGTCTACCTGCTACAGGTGAAAAATT 840  
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
Db 841 ACAGATATCGAAATCCACCCAGACCGTATGCTAGCCGTATCTCCCTTGGCAGGGGAT 900  
QY 301 ValLeuSerIleGluLysAlaGlnInAspValAspGlnGluLysAlaLysAspLeu 320  
Db 901 ATGCTCATTGATGAGAAAGCTTCACGGAATACGATGACAAAGAACCCCTTGAATG 960  
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
Db 961 GCTGAGAAGATCGCGAAACACCTTTGATTTAATGATTTCATCGATCAATTAGATCAG 1020  
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
Db 1021 GTGCAAAATATGGGCGCATGGAAGACTTGCTCAAGATGATCCAGGTATGCCCAACAAT 1080  
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
Db 1081 CCAGCACTTCAAAACATGAAGTGGATGAACGCCAGATTGCTGTAAACGTGCCATTGTG 1140  
QY 381 GlnSerMetThrProAlaGluArgAsnAspProAspThrLeuAsnValSerArgLysLys 400  
Db 1141 TCTTCGATCACATCTGGAAGACGTGAAACCCACAGATTGTTAATCCAAAGCGCTCGCGT 1200  
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
Db 1201 CGTATTGCTGCTGGTTCTGGAATATACATTCGGAAGTCAATAAATTCATCAAGAGACTTT 1260  
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys 440  
Db 1261 AACCAGCTTAAACAGCTCATCGAGGGTGTATGCTCGGG----- 1299  
QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMet-----AsnLeuPro 454  
Db 1300 -----GATATGAATAAATATGATGAAGAAATGGGATTAATCCAAATACCTTCT 1350  
RESULT 11  
US-09-738-626-2255  
; Sequence 2255, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOUCUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 2255  
; LENGTH: 1641  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-2255  
Alignment Scores:  
Pred. No.: 5,84e-122 Length: 1641  
Score: 1168.00 Matches: 237  
Percent Similarity: 66.67% Conservative: 85  
Best Local Similarity: 49.07% Mismatches: 125  
Query Match: 51.36% Indels: 36  
DB: 5 Gaps: 5  
US-09-943-108a-2 (1-455) x US-09-738-626-2255 (1-1641)  
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22  
Db 4 TTTGAGTCACTCTCGGATCGGTTGATAGCGCGCTTTCGGCGCTTCGGCGCAAGGAAAG 63  
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42  
Db 64 CTCACCGAGCGAGACATCAATGCAACACACGCGAGATCCGCTCGCGCTGCTGGAAGT 123  
QY 43 AspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62  
Db 124 GAGCTTCATTAAACGGTTGTCGTTCCTTCAATTAACCGAATCAAGAACGCGCGTGGT 183  
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82  
Db 184 GCAGAAGTTCTCAGGACACTCAACCGCGCGCAAGTCAATCAAGATCGTCAACGAGAA 243  
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102  
Db 244 CTGTTTCAGATCTCGGTCGGGAAACCCCGCGACTGTCTACTGGCCAAACCCACGACC 303  
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122  
Db 304 GTCATCATGCTGGAGGTCCTCAGGGTGCAGGTAGACACCCCTCGCAGGTAACTGTCC 363  
QY 123 LeuLeuMetArgLysTyAsnLysLysProMetLeuValAlaAlaAspIleTyrArg 142  
Db 364 AAGCACTCGTCAAGCAG---GGTCACACTCTCTCTGTTGCTGTGACCTTCAGCGT 420  
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyr--- 161  
Db 421 CCAGGCGCAGTTTCAGCAGCTGCAAAATTTGGGTGAACGCGCAGGCGTACCACTTCGCA 480  
QY 162 -----SerGluGlyAsp 165  
Db 481 CCGGATCCAGGCAACCAAGCATCGATCCCTCGAGCAGAAATGGGCACCTCCCAAGGTGAT 540

```
QY 166 GlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeu 185
Db 541 -----CGATCGAGGTAGCGCGCGAGGTATCGAAGAACCGACGCGACCGACGAC 591
QY 186 AspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGlu 205
Db 592 GACATCGATGATCGGATACCGAGGTGCGCTGCGTATCGATGAAACCGCTGATGACTCAG 651
QY 206 LeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSerMet 225
Db 652 GCACGCAACATCCGCGAAGCCATCAACCCCTGATGAAGTCTCTTTCTGATTCATGCTCATG 711
QY 226 ThrGlyGlnAspAlaValAlaGluSerPheAspAspGlnLeuAspValThrGly 245
Db 712 ATTGTCAGAGCGCTAGACACCGCGAGAGATTCGCGACGCGCTGACCTTCCCGGT 771
QY 246 ValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAlaLeuSerIleArgSer 265
Db 772 GTTGTCCTGACCAAGCTTGATGGCGACGCGCGCGTGTGCTGCACATATCCATCCGCGAA 831
QY 266 ValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeuAspGlyLeuGluLeu 285
Db 832 GTCACCGCGACGCGCATGATGTTGCTCCACTGCTGAAACACTCCAGCACTTCGACGTC 891
QY 286 PheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIle 305
Db 892 TTCACCCAGAGCGCATGGCCAGCGAATCTCTGGCATGGGTGACGTACTGTCATCATC 951
QY 306 GluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuLysLysMetArg 325
Db 952 GAGCAGCGCGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1011
QY 326 GluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGlnValLysAsnLeuGly 345
Db 1012 TCGCGCGAGCTACCCCTGGAGACTTCTTACCAAAATGCTGATGATGATGATGATGATG 1071
QY 346 ProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLys 365
Db 1072 CCAATCGCAACATCCTCAAGATGCTGCTGTGCGAAGCAGATGTCCTCAATGCGCGAC 1131
QY 366 LeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLeuIleGlnSerMetThrPro 385
Db 1132 ATG--GTTGATGAGAGCAACTCGACCGCATCCAGCGATATTCGCGCGGTATGACCCG 1188
QY 386 AlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLysArgIleAlaLysGly 405
Db 1189 GCGGAGCGGATATCCAAAGATCTCAACGCTTCCAGGCGCAGCGCATCCGCCAACCGGT 1248
QY 406 SerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPheAsnAspMetLysLys 425
Db 1249 TCCGCTGTGACCGTGTCCGAAGTAAACAACTGTGTGAACGCTTCTTCGAGGCTCGCAAG 1308
QY 426 MetMetLysGlnPheThrGly----- 432
Db 1309 ATGATGGGTCAAAATGGCTGGCATGGGTCTCGATCCCGGACGATCCACCAAG 1368
QY 433 ---GlyGlyLysGlyLysLysGlyArgAsnGlnMetGlnAsnMetLeuLysGlyMet 451
Db 1369 AGCAGACCGAGCGCGCGAGGATAGAACGCGCAGCGTAAACCGACCGAAGAGGCGCCA 1428
QY 452 AsnLeuPro 454
Db 1429 ACCAGCCA 1437
RESULT 12
US-09-815-242-6166
; Sequence 6166, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITEA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6166
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; LOCATION: (1)...(1362)
US-09-815-242-6166
Alignment Scores:
Pred. No.: 1,11e-121 Length: 1362
Score: 1164.50 Matches: 231
Percent Similarity: 71.74% Conservative: 94
Best Local Similarity: 50.99% Mismatches: 121
Query Match: 51.21% Indels: 7
DB: 10 Gaps: 2
US-09-943-108A-2 (1-455) x US-09-815-242-6166 (1-1362)
```

```
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTTGATATTTAAACGATCGTTTTCGCGCACGCTGGCGCAATATCATGTCGCGTGGAGCG 63
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
Db 64 CTCACCTGAAGACACGTAAGAATACGCTGCGGAGTGGCATGCGCTGCTGGAGCG 123
QY 43 AspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db 124 GACGTAGCTCTCGCGTGTGTCGCTGAGTTTATCAATCGCTAAAGAGAAAGCGGTGCT 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
Db 184 CATGAAGTTAATAGAGCCCTGACGCGGCGAGAGTTCGTCATAATAGTCCGTACCGAA 243
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
Db 244 CTGGTTGCGCGATGGCGAAGAGAACCCAGACCCCTGAACCTGCTGCGCAACCGCTCG 303
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTCGTACTGATGCGGCGCTGCAAGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 363
QY 123 LeuLeuMetArgLysLysTyrosinLysProMetLeuValAlaAlaAspIleTyArg 142
Db 364 AAGTTCCTGCGCGAAGACCAAGAGAAAGTGTGTTGTTCTTCCGACCGCTTATCGC 423
```

```
QY 143 ProLaalaleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 424 CCGCGCAATCAACAGCTTGAGAGCGTGGCAGAGAGTGGCGGTGATTTCTCCCT 483

QY 163 GluGlyaspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 484 TCTGATGTTGTCAGAGCGGTAGATATCGTTAAACGCGCGTGAAGAAGCCAACTG 543

QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyValGlnHisIleAspLysAlaLeu 202
...  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 544 AAATCTACAGCGTGTCTGTGTGATACCGTGTCTGTCTGTCTGTCTGTCTGTCTGT 603

QY 203 MetAsnGluLeuLysGluValLysGlnIleAlaLysProAsnGlnIleMetLeuValVal 222
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 604 ATGGACGAGATCAACAGTCCATCGTCGATTACCGGTTCAACCCCTGTTGTGTT 663

QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGlnIleSerPheAspGlnLeuAsp 242
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 664 GACGCCATGACCGGTGAGTACGCGCAATACGCGCAAAAGCATTCATGAAGCGTTACCG 723

QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyValSerThrArgGlyValAlaLeuSer 262
...  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 724 CTTACCGCGTATGTTGACCAAGTGACGCGCATGCCGCGCGGTGCGCGCTCTCT 783

QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 784 ATTGTCACATCACTGGCAACCGATCAAGTCTCTCGGTGTTGGCGAGAAACTGAGCGG 843

QY 283 LeuGlnLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 844 CTGGAGCGGTTCATCCGACCGCATCGCGTCGCTATTCCTCGGCATGGCGAGCTACTG 903

QY 303 SerLeuIleGlnLysAlaGlnIleAspValAspGlnIleLysAlaLysLeuGlnLys 322
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 904 TCGCTGATCAAGATATCAAGAAAGATTGACGCGCGCGCGAGAGAAATAGCCAGC 963

QY 323 LysMetArgGlu---SerSerPheThrLeuAspPheLeuGluGlnLeuAspGlnVal 341
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 964 AAGCTGAAAAAGGTGACGGCTTCGATCTCAACGACTTCTTGAGCAGCTGCGCCAGATG 1023

QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLys 361
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1024 AAAAATATGGCGGCATGCTAGTCTGATGGCAAGCTGCGCGCATGGCGCAGATCCCG 1083

QY 362 GlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLeuGln 381
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1084 GATAACGTCAAGTCAAGATGACGATAAAGTGTGTCGATGGAAGCCATCATCAAC 1143

QY 382 SerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLysArg 401
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1144 TCGATGACGATGAAGAGCGCGCTAAGCCAGAAATCATCAAGGTTCCGCTAAGCGCGT 1203

QY 402 IleAlaLysGlySerGlyArgSerLeuGlnValAsnArgLeuMetLysGlnPheAsn 421
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1204 ATTCTCCGGTTTCGGGTATGACGATGACGATAAAGTGTGTCGATGGAAGCCATCATCAAC 1263

QY 422 AspMetLysLysMetMetLysGlnPheThrGlyGlyLysLysLysLysLysArg 441
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1264 GACATGACGCGATGATGAG-----AAATGAAGAGGCGCGGAATG 1305

QY 442 AsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuPro 454
...  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1306 CGGAGATGATGAGAGCATGAGGCTATGATGCCCCCA 1344
```

## RESULT 13

```
US-09-815-242-1949/c
; Sequence 1949, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 1949
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-815-242-1949
```

```
Alignment Scores:
Pred. No.: 2,4e-121 Length: 721
Score: 1158.00 Matches: 234
Percent Similarity: 99.58% Conservative: 2
Best Local Similarity: 98.73% Mismatches: 1
Query Match: 50.92% Indels: 0
DB: 10 Gaps: 0
```

US-09-943-108a-2 (1-455) x US-09-815-242-1949 (1-721)

```
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 713 ATGGCATTTGAAGGTTATCAGAACGCTTGCACGCGATGCAAAAAATGCGTGTGAAG 654

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 653 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTACTT 594

QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 593 GAGCGTACGCTAAACCTTAAAGTGAAGAATTTATTAAACAGTATCAGAACGCGCA 534

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGln 80
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 533 TTAGTTCCGATGTAATGCAATCATTAACACAGCGCAACAAGTTATTAAATAGTTCAA 474

QY 81 AspGlnLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 473 GATGAATTAACGAGTGTGTTGAGGAGAAAAATACGTCGATTAATATGTCATAAATACCA 414

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyThrThrAlaGlyLys 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 413 CCTACTGTGTTATGATGTTGTTTACAGGTGCTGTAACAAACACAACTGCGAGTAAA 354

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 353 TTAGCATTTATGATGCTAAAAAATACACAAAAACCTATGTTAGTTCGACAGATATT 294

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 293 TATCGTCCAGCAGCGATAAATCAATACAAACAGTAGGGAACAAATGATATCTCTGTA 234
```



```
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 233 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGAATTAATGCAATTAACATGCT 174
QY 181 LysGluGluHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 173 AANGAAGAACATTTAGACTTTGTAATCATGTATACAGAGGTGCAATTAACATCGATGAA 114
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 113 GCATTGATGAGAGATTAAGAGAGTAAGAGACATGCTAAACCAAGCAATATATGTA 54
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPhe 237
Db 53 GTTGTGCAATCAATGACGGGTCAAGATGCTGTCAATGTTGAGAAATCTTTT 3
RESULT 14
US-09-815-242-7850
; Sequence 7850, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7850
; LENGTH: 1374
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1374)
US-09-815-242-7850
Alignment Scores:
Pred. No.: 4,79e-120 Length: 1374
Score: 1150.00 Matches: 231
Percent Similarity: 70.90% Conservative: 93
Best Local Similarity: 50.55% Mismatches: 121
Query Match: 50.57% Indels: 12
DB: 10 Gaps: 3
US-09-943-108a-2 (1-455) x US-09-815-242-7850 (1-1374)
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTGCAAAACCTTACAGACCGGCTCTCGCAGACGCTTCGCCATGTACCGGCAAGGCCAAG 63
```

```
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
Db 64 CTGACCGAGGACAAACATCAAGGACACTCTGCGGAGAGTGGCATGGCCCTGCTCGAGGCC 123
QY 43 AspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db 124 GAGCTGGCCCTCGCGGTGGTCAAGGACTTCGTCAACAAGGTCAAGAGAGCGCGCTCGGT 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
Db 184 ACCAGGCTCTGAGAGCGCTGACCCGGACAGCGGCTTCGTGAAGATCGTCGCCGCCGAG 243
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProProThr 102
Db 244 CTGAGGAGCTGATGGGGGGCGGCAACGAGGACCTGGCGCTGAGCGTCGCGCGCGCGG 303
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTGATCTGATGGCGGCTGCGAGGGCGGGCAAGACACCCCGGGCAAGCTGGGCG 363
QY 123 LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArg 142
Db 364 CGTTTCCTTAAGGACCGCAAGAGAGTGGTGATGGTGGTTCGCCGACGCTACCGA 423
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
Db 424 CCGCGCGCTATCAAGCAGCTGGAACCCCTGGCGGGCGAGGTGCGCGTAACCTTCTCCT 483
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
Db 484 TCGAGGCTCAGCGCAAGCGGTTGGCCATCGCGGAGCGGCGATCCGCGAGCGGAGGTG 543
QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu 202
Db 544 AATTCATCGAGTGGTGTGCTGATCGATACCGAGCGCGCTGACATCGATGCGCGCATG 603
QY 203 MetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValVal 222
Db 604 ATGAGCAGATCAAGCAGGTGCACGGCGGCGATCAAGCGCGGGAACCCCTGTTCTGTC 663
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
Db 664 GATGCCATGACCGCGCAGATGCGCCCAACCGCCCAAGGCTTCATGACGCTTGGCGG 723
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 724 CTGACCGCGGTGGTCTGACCAAGTGCAGCGGCGCGCTGGTGGTGGCGGCTCTCG 783
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
Db 784 GTCGGGCGATCACCGCAAGCGGATCAAGTTCTCGGCATGGCGGCAAGAGCGAAGCG 843
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 844 CTGATCCGTTCCATFCCCGACCGGTGGCTCGCATCCCTCGGCATGGCGGCGGCTGTC 903
QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGluLys 322
Db 904 AGCTTCATCGACAGCGCGAGCAACCTCGACCGCGCAAGCGCGAAGCTGGCGAAG 963
QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnVal 341
Db 964 AAGATCAGAAGGCAAGGGCTTCGACCTCGAAGAGATTCCTCGCGCACCAGTCAACAGATG 1023
QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIlePro-----GlyMetAsn 358
Db 1024 AAGACATGGCGGCGCTCGCGGGGCTCATGGACAAAGCTGCGCGATGCTCGCGGGGTCAAC 1083
QY 359 LysMetLysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAla 378
Db 1084 CTGGCCGAGATGGGCAATGCCCCAGGTGCTGGGAGAAACAGTTCACAGCAGATGGAGGCG 1143
QY 379 IleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArg 398
```

```
Db 1144 ATCATCAACTCGATGACGCGCGCAAGCGCGATCCGGAATGATCAGCGGCTCGCG 1203
QY 399 LysLysArgIleAlaLysSerGlyArgSerLeuGlnGluValAsnArgLeuMetLys 418
Db 1204 AAGCGCGCATCGCCTTGGCTCGGTACCCAGGTGCAGAGCGTCGCGCGCTGATCAAG 1263
QY 419 GlnPheAsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLys 438
Db 1264 CAGCACAGCAGATGCAGAGATGATGAGAAGGTACCAGCAAGGCG----- 1311
QY 439 GlyLysArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1312 -----GGCATGCCCAAGATGATGCGAGCATGGCAGTATGTTTC 1350
```

RESULT 15

```
US-09-815-242-6868
; Sequence 6868, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamanoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6868
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (1)...(1389)
US-09-815-242-6868
```

Alignment Scores:

```
Pred. No.: 1,09e-118 Length: 1389
Score: 1138.00 Matches: 225
Percent Similarity: 70.67% Conservative: 93
Best Local Similarity: 50.00% Mismatches: 122
Query Match: 50.04% Indels: 10
DB: Gaps: 2
```

US-09-943-108a-2 (1-455) x US-09-815-242-6868 (1-1389)

```
QY 3 PheGluGlyLeuSerGluArgGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTGAGATTTATTCGATCGCCTTCCAAACCTAGCTAATATCACAGGAAGGCGGT 63
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
```

```
Db 64 TTAACGGAAGATAATTAAGAAACCTTACGCGAAGTCGCTATGCGATTACTTGAAGCC 123
QY 43 AspValAsnPhelYsValValYsGluPheIleYsThrValSerGluArgAlaLeuGly 62
Db 124 GATGTTCCCTTGCCTGTCGTGCGTGAATTTATGCAAAAAGTAAAGAAAGCGCGCTTGGG 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleYsIleValGlnAspGlu 82
Db 184 GAAGAGTCANTAAAGATTTACGCCAGGCAAGAAATCTTAAANAATCGTTCACGCTGAG 243
QY 83 LeuThrLysLeuMetGlyGlyGlnAsnThrSerIleAsnMetSerAsnLysProProThr 102
Db 244 CTTGAAAAACCATGGCGAAGCAATGAGAGTTTAAATCTCGCAACCCACCACCAGCA 303
QY 103 ValValMetMetValGlyLeuGlnGlnAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTTATCTTAATGGCGGTTTACAAGGGCGGGTAAACACCAGTGGGTAAATTGGCA 363
QY 123 LeuLeuMetArgLysLysTyraAsnLysLysProMetLeuValAlaAlaAspIleTyraG 142
Db 364 AAATCTTGGGTGAAGCCATAAAGAAAGAGTGTAGTGTCTCTGCTGACGTATATCGC 423
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTySer 162
Db 424 CCTGCTCGGATTAAAGCAACTTGAACACCTTGGCTCAATCCGTTGGCGTGGATTTTTC 483
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
Db 484 TCGGATGTTAAACAAACCCCGTTGATATTGCTAAATCGCGCTTGCCTGCTCAAAACTG 543
QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu 202
Db 544 AAATCTACGATGTTGTTGTTGGTACGCGAGTTCGCTACACGTTGATACAGAATG 603
QY 203 MetAsnGluLeuLysGluValYsLysIleAlaLysProAsnGlnIleMetLeuValVal 222
Db 604 ATGGACGAAATCAAGCAAGTCATTCATTAATCAATCAATCAATCAATCAATCAATGTT 663
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGluLeuAsp 242
Db 664 GATCGATGATCGTTCAGATGCGGCAAAATACAGCAAAAGCCCTTAAATGAAGCATTCCT 723
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 724 CTTACAGGGGTTATTTGACGAAAGTGCAGCGGTGATCGCGCGCGTGGTGGCGCTTATCG 783
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
Db 784 ATTCGCAAAATCACAGGTAAACCAATCAATCAATCTTGGGTGTGGCGCGAAGAACAGAGCG 843
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 844 CTTGAGCCATTCCTCTGATCGTGTGCTTCCCGTATTTTGGGCATGGCGGATGCTT 903
QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeuLys 322
Db 904 TCCCTTATCGAAGATCTTGACGTTCTGTGTGATGCTGAAAGAGCGGAAAAATGGCGCAG 963
QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnVal 341
Db 964 AAATTCAGAAAGAGCGGATGATTACTTACTTATGATGATTTCGCGGCAACAGCTGATCGA 1023
QY 342 LysAsnLeuGlyProLeuAspAspIleMetIysMetIleProGlyMetAsnLysMetLys 361
Db 1024 AAAAAATGGCGCGCATGATGCTATGCTGAGAAATTTACCAAGTGCAGAAATTTCTCT 1083
QY 362 GlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIleGln 381
Db 1084 GAGCACGCTTAAATCAAGTAGATGACAAATGTTTGTCAAAATGGAAGCGATCAATTAC 1143
QY 382 SerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgGlyLysArg 401
```



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 01:10:37 ; Search time 1563 Seconds  
(without alignments)  
4714.617 Million cell updates/sec

Title: US-09-943-108a-2  
Perfect score: 2274  
Sequence: 1 MAFEGLSERLQATMKMRGK.....GKKGRNQNMKGMNLPF 455

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet\_p2n.model -DEV=rlp  
-Q=/cgn2\_i/USPTO\_spool/US09943108/runat\_21022003\_151631\_1824/app\_query.fasta\_1.647  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09943108 -CGN\_1\_1\_2874 @runat\_21022003\_151631\_1824 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_estl:\*
  - 10: gb\_est2:\*
  - 11: gb\_htc:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vrt:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	1074	47.2	2169	17	BH770912 LIMGtag64
c 2	757.5	33.3	875	17	AF075878 AF075878
c 3	715.5	31.5	773	13	B1934118 EST594007
c 4	631	27.7	656	14	B0805434 WHE3566.H
c 5	630	27.7	711	12	B940797 7039A06.X
c 6	616.5	27.1	2387	11	AK011928 Mus muscu
c 7	594.5	26.1	688	17	BH394904 AG-ND-138
c 8	562.5	24.7	605	10	BE432304 EST398833
c 9	558	24.5	593	10	AW648731 EST327101
c 10	555.5	24.4	581	9	AI977909 496035E06
c 11	534	23.5	677	13	BM413418 EST587745
c 12	520.5	22.9	1200	11	AY109321 Zea mays
c 13	513.5	22.6	533	10	AW930742 EST336585
c 14	506.5	22.3	677	14	BQ996389 QGG12L07.
c 15	485	21.3	514	13	BM323434 PIC1_19.C
c 16	479	21.1	1020	17	CNS07B0K AL437274 T3 end of
c 17	473	20.8	791	17	AQ32477 nbxb0041K
c 18	465.5	20.5	1032	11	AY106667 Zea mays
c 19	463.5	20.4	483	10	BE596266 P11_51.F0
c 20	451	19.8	777	17	BH402176 AG-ND-122
c 21	448.5	19.7	490	13	BM325805 PIC1_53.A
c 22	438.5	19.3	554	17	BH369841 AG-ND-126
c 23	437.5	19.2	481	17	BH376496 AG-ND-178
c 24	433.5	19.1	880	17	CNS077VW AL433218 T7 end of
c 25	420	18.5	465	9	AI087737 SNAMC23
c 26	419	18.4	476	10	AW033339 EST276910
c 27	419	18.4	899	14	BQ736950 AGENCOURT
c 28	414.5	18.2	547	10	BE461473 EST412892
c 29	412.5	18.1	608	13	B1933601 EST553490
c 30	412.5	18.1	1108	9	AU091185 AU091185
c 31	408	17.9	1059	14	BM806783 AGENCOURT
c 32	405	17.8	1012	17	AF094895 AF094895
c 33	400	17.6	949	17	CNS076TA AL431828 T7 end of
c 34	398.5	17.5	814	17	AZ529841 ENTC084TR
c 35	392.5	17.3	889	17	AZ682189 ENTCB30TF
c 36	389.5	17.1	773	13	B1454259 603170687
c 37	388	17.1	761	9	AL655153 AL655153
c 38	385.5	17.0	656	12	BF641333 NF064F12I
c 39	380.5	16.7	694	13	BJ473333 BJ473333
c 40	378.5	16.6	874	12	BG679799 602626480
c 41	377	16.6	707	17	AQ654845 Sheared D
c 42	375.5	16.5	753	13	BG921695 602825514
c 43	375	16.5	382	10	AM649149 EST327603
c 44	374.5	16.5	975	13	BG919869 60283428
c 45	369.5	16.2	573	12	BG317280 947027E11

ALIGNMENTS

RESULT 1  
BH770912/c  
LOCUS BH770912 2169 bp DNA linear GSS 01-MAY-2002  
DEFINITION LIMGtag644 MG1363 Random Sequence Tag Library Lactococcus lactis  
subsp. cremoris genomic, DNA sequence.  
ACCESSION BH770912.1 GI:20373869  
VERSION BH770912.1  
KEYWORDS GSS.  
SOURCE Lactococcus lactis subsp. cremoris  
ORGANISM Lactococcus lactis subsp. cremoris  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
REFERENCE 1 (bases 1 to 2169)  
AUTHORS Bolotin,A., Ehrlich,S.D. and Sorokin,A.  
TITLE Studies of genomes of dairy bacteria Lactococcus lactis

[illegible]

QY	275	GlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArg	294
Db	1568	GGGACTGGTGA AAAATTAACCGACCTTTTGAAGTTTTTACCGTGAATCGCTGATGAGCTCAAGA	1509
QY	295	IleLeuGlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnAspValAspGln	314
Db	1508	ATTCTTGGTATGGGTGATATGCTCACTCTGATTGAGAGGCCACGACCTAATATGATGAG	1449
QY	315	GluLysAlaLysAspLeuGluLysLysMetArgGluSerSerPheThrLeuAspAspPhe	334
Db	1448	GAACAATCTGC AAAACTCGCTGAA AAAATGCGTGAATACCGTTTTCATTATGAGGATTC	1389
QY	335	LeuGluGlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIle	354
Db	1388	CTTGAGCAACTAGATCAAGTTACAAATATGGGACCAATGGAAGATATTATGAAAATGATT	1329
QY	355	ProGlyMetAsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAsp	374
Db	1328	CCAGATATGTC ACAAATGCCCTGGCTTGGATTAAGTTAAAGTTGACCCCTAAAGATGTGCT	1269
QY	375	HisIleLysAlaIleIleGlnSerMetThrProIaGluArgAsnAsnProAspThrLeu	394
Db	1268	AGAAAACGACCAATGTTATGCAATGACCCAGCTGAACCTCAITTTAGAAGCAGAACTA	1209
QY	395	AsnValSerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsn	414
Db	1208	TGGCAGCTCGTCGTCGCAATCGCCGAGTCTCGGAATTCATTATTGGAAGTAAT	1149
QY	415	ArgLeuMetLysGlnPheAsn	422
Db	1148	AAATTCATTATGCAATTTATCAATCCAAAGAATATGCAAGGAATCATGAATGTGTAT	1099
QY	423	MetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLysArgAsn	442
Db	1088	ATGAATCGCATGATCAAAAATATGATGGGTGGAGCTGGTGGG	1047
QY	443	GlnMetGlnAsnMetLeuLysGlyMetAsnLeuPro	454
Db	1046	CAATGCCAATATGCCAACCGGTTCTGGAATGCCA	1011
RESULT 2	AF075878	875 bp DNA linear	GSS 29- AUG
LOCUS	AF075878	salmonella typhimurium LT2, Lambda DASH II	Salmonella
DEFINITION	AF075878	typhimurium genomic clone 245-r3, DNA sequence.	
ACCESSION	AF075878		
KEYWORDS	AF075878.1	Gi:3320748	
SOURCE		GSS.	
ORGANISM		Salmonella typhimurium.	
REFERENCE		1 (bases 1 to 875)	
AUTHORS		Wong, R.M.-Y., Wong, K.K., Benson, N.R. and McClelland, M.	
TITLE		Sample sequencing of a Salmonella typhimurium LT2 lambda library	
JOURNAL		comparison to the Escherichia coli K12 genome	
MEDLINE		PMWS Microbiol. Lett. 173 (2), 411-423 (1999)	
COMMENT		9243757	
		Contact: McClelland M	
		Molecular Biology	
		Sidney Kimmel Cancer Center	
		3099 Science Park Road, San Diego, CA 92121, USA	
		Email: mclelland@lifsci.sdsu.edu	
		Class: shotgun.	
FEATURES		Location/Qualifiers	
source		1..875	
		/organism="Salmonella typhimurium"	
		/strain="LT2"	
		/db_xref="taxon:602"	
		/clone="245-r3"	
		/note="vector: Lambda DASH II; sequenced using Li-Cor	
		/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"	

```

BASE COUNT      178 a  266 c  217 g  214 t
ORIGIN
sequencer"
Alignment Scores:
Pred. No.:      1,51e-73      Length:      875
Score:          757.50      Matches:      152
Percent Similarity: 73.08%      Conservative: 57
Best Local Similarity: 53.15%      Mismatches: 76
Query Match:    33.31%      Indels:      2
DB:             17      Gaps:      1
US-09-943-108a-2 (1-455) x AF075878 (1-875)
QY 87 MetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProProThrValValMetMet 106
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 867 ATGGCGAAGAGAACCGAGCGCTGAATTTTAC-TGCGACGCGCGCGAGCTAGTATGATG 809
QY 107 ValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArg 126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 808 GCGGCTGCGAGGGGGGGTAACACACAGCGTCGTAAGTTGGTAAATCTTGTGGC 749
QY 127 LysLysTyAsnLysLysProMetLeuValAlaAlaAspIleTyArgProAlaAlaLe 146
    :|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
Db 748 GAGAGCACAGAGAGAGAGTCTGCTGCTGCGACGCTATCGCGCGGGCGGATC 689
QY 147 AsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTySerGlnGlyaspGln 166
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 688 AAACAGCTCGAAGCGTGGCTGACGAGGTGGCGGATTTCTTCGCTGTGATGTCGCG 629
QY 167 ValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlnGluHisLeuasp 186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 628 CAGAAACCGGTGATGCTGCTCAGCGCGCGCTGAAGAGCGAGCTCAATCTACGAC 569
QY 187 PheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeu 206
    :|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
Db 568 GTGCTGCTGGTGATACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 509
QY 207 LysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValaspSerMetThr 226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 AAACAGGTCCAGCTTCTATCAGCCAGTAGAAGCGCTGTTGCTGCTGCTGCTGCTGCTG 449
QY 227 GlyGlnAspAlaValAsnValAlaGlyLysPheAspAspGlnLeuLysValThrGlyVal 246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 GGTGCGATGCGCGGAATACCGCAAAAGCCTTTACGAGAGCGCTGCGGTTAACCGCGTG 389
QY 247 ThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArgSerVal 266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 GTGCTGACCAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 329
QY 267 ThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPhe 286
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 328 ACCGGCAACCGGATTAATTCCTCGGTGTCGGCGAGAAACCGAGCGCTGGAGCCCATTC 269
QY 287 HisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIleGlu 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 CATCCGGATCGATCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 209
QY 307 LysAlaGlnGlnAspValaspGlnGlnLysAlaLysAspLeuLysLysMetArgGlu 326
    :|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
Db 208 GATATCAAGACCAAGTTGACCGCGACAGCTGAGAGCTGGCGACCAACTGAAGAA 149
QY 327 ---SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGly 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 GCGGACGTTTCGACCTGAACGACTCTCTGGAACAGCTCAAAACAGATGAATAACATGGG 89
QY 346 ProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLys 365
    :|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
Db 88 GGTATGCCCATGCTGATGGGCAAAATTACCGGCGATGGGCGCATGTCGGCAACAGTTAAA 29
QY 366 LeuAsnMetSerGlnLys 371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 28 TCGCAGATGGATGACAG 11

```

## RESULT 3

BI934118

LOCUS

DEFINITION

EST554007 tomato flower, anthesis Lycopersicon esculentum cDNA

clone cTOD18T16 5' end, mRNA sequence.

ACCESSION

BI934118

VERSION

BI934118.1

KEYWORDS

EST

SOURCE

ORGANISM

tomato.

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 773)

van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,

Utterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,

Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

Generation of ESTs from tomato flower tissue, anthesis (2001)

Unpublished (2001)

Contact: CUGI

Clemson University

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics

Institute

Seq primer: T3.

Location/Qualifiers

1..773

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone="cTOD18T16"

/clone\_lib="tomato flower, anthesis"

/tissue\_type="flower"

/dev\_stage="anthesis"

/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; supplier: Cornell University; sequencing: The

Institute for Genomic Research; Flower buds and flowers

were taken from greenhouse plants (4-8 wks old, TA496).

They were immediately frozen in liquid nitrogen and then

size-separated while remaining frozen."

BASE COUNT 227 a 129 c 210 g 207 t

ORIGIN

Alignment Scores:

Pred. No.: 5,83e-69 Length: 773

Score: 715.50 Matches: 134

Percent Similarity: 75.19% Conservative: 60

Best Local Similarity: 51.94% Mismatches: 63

Query Match: 31.46% Indels: 1

DB: 13 Gaps: 1

US-09-943-108a-2 (1-455) x BI934118 (1-773)

QY 39 LeuPheGluAlaaspValasnPheLysValLysGluPheIleLysThrValSerGlu 58

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1 CTTTATGAAGCTGATGTAGTCTCCCAAGGTTTGTTCAGTCTGTAGTGA 60

QY 59 ArgAlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnValIleLysIle 78

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 61 GAAGCCGTGGCAGCTGGTGTGATTCGAGGTGTAACACAGATCAGCACTAGTAACT 120

QY 79 ValGlnAspGlnLeuThrLysLeuMetGlyGlyLysThrSerIleAsnMetSerAsn 98

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 121 GTAGTGACGACGCTGTGAACCTGATGGGTGGAGAGGTTCTGAACTGGTTTGTCTAAA 180

QY 99 LysProProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAla 118

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 181 TCTAAGCCCAACCGTAATACTATTGGCGGCTCTACAAGGTGTTGGAGACAACTGTAGT 240

```

QY 119 GlyLysLeuAlaLeuMetArgLysLysTyrAspLysProMetLeuValAlaAla 138
Db 241 GCRAAGTACCTTTATCTA---AAGACGAGGTAAAGTGCATCTGCTGCTGGA 297
QY 139 AspIleYrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIle 158
Db 298 GACGTGTACAGACCTGCTGCTATTGACCACTTGTTATTTGGTAAACAGGTTGATGA 357
QY 159 ProValTySereGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLys 178
Db 358 CCTTTATGACAGCAAGACAGATGTAACCTGCAGAAATAGCCCGACAGGATTACAA 417
QY 179 HisAlaLysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIle 198
Db 418 GAGCCAAAAGAGAGATGCTAGTCTATATGATGATGACGCTGGAGACTTCAGATA 477
QY 199 AspGluAlaLeuMetAsnGlnLeuLysGluValLysGlnIleAlaLysProAsnGlnIle 218
Db 478 GATAAACTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537
QY 219 MetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAsp 238
Db 538 TTACTTTGTCGACGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
QY 239 AspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGly 258
Db 598 CTCGAAATTTGGGATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657
QY 259 AlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlu 278
Db 658 GCACGCTTTTGTCTCAGAGGATGCTCAGGAAAGCCAACTCAAGCTGCTAGGAGGGTGA 717
QY 279 LysLeuAspGlyLeuGlnLysPheHisProGluArgMetAlaSerArgIleLeu 296
Db 718 CGTATGAGGACCTTGAACCTTTCTAUCCTGACCGCATGCTGGACGATTTTITA 771

RESULT 4
BQ0805434
LOCUS
DEFINITION WHE3566_H12_P24ZS Wheat developing grains cdna library Triticum
aestivum cDNA clone WHE3566_H12_P24, mRNA sequence.
ACCESSION BQ0805434
VERSION BQ0805434.1 GI:22029643
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 656)
AUTHORS Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin
X., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J.,
Wilson,C. and Woo,J.
The structure and function of the expressed portion of the wheat
genomes - Developing grains cdna library
Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES
Location/Qualifiers
1..656
/organism="Triticum aestivum"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3566_H12_P24"
/clone_lib="Wheat developing grains cdna library"

```

```

/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/Note="Vector: Lambda ZAP II, excised phagemid; Site_1:
EcoRI; Plants were grown under six following different
environmental regimes in greenhouse, Environment 1)
240C/170C day/night, well-watered, with post-anthesis
fertilizer, Environment 2) 240C/170C day/night,
well-watered, without post-anthesis fertilizer,
Environment 3) 370C/170C day/night, well-watered, with
post-anthesis fertilizer, Environment 4) 370C/170C
day/night, well-watered, without post-anthesis fertilizer,
Environment 5) 370C/170C day/night plus drought, with
post-anthesis fertilizer, Environment 6) 370C/170C
day/night plus drought, without post-anthesis fertilizer,
developing wheat grains from the following were excised
and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28,
32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24
, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16,
20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16
, 20, 24, 28, 30 DPA and total RNA was prepared by S.
Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA
library was made using poly (A) RNA, and the cDNA clones
were in vivo excised to give plasmid SK(-) phagemids
in the TJ Close lab (Chin, Close, Fenton) at the
University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (others)."
BASE COUNT 200 a 115 c 179 g 162 t
ORIGIN
Alignment Scores:
Pred. No.: 11e-59 Length: 656
Score: 631.00 Matches: 116
Percent Similarity: 75.23% Conservative: 48
Best Local Similarity: 53.21% Mismatches: 54
Query Match: 27.75% Indels: 0
DB: 14 Gaps: 0
US-09-943-108a-2 (1-455) x BQ0805434 (1-656)
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 3 TACAGGCTCTGCCATGCTCACTCTACTGCTGGTGACAGGTGGGTGCCAGTT 62
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 63 TACTCAGAGGAACCGCGCCCAACCTGCAGAAATACCAAGAATCCCGTGAAGAGCG 122
QY 181 LysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 123 AAAAGAAATAATATATGTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 182
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGlnIleAlaLysProAsnGlnIleMetIleu 220
Db 183 ACAATGATGTTGAATTAAGAAAGATGAAGAGCCAGTTAATCTTACAGAGATTTCCTT 242
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 243 GTTGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 303 ATTGTGATATCCGCTGCTATATTGCTAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 362
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysIleu 280
Db 363 CTTAGTGTGAAGAGGTGCTGGAAGCCCATCAAGTTTCTTGGCGGAGGAGGAGGATG 422
QY 281 AspGlyLeuGlnLysPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300

```

```

Db 423 GAGGACCTTGAGCTTTCATCCGACCGCATGCACACAGTGTTCGGATGGAGAT 482
    ::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 ValLeuSerLeuLeuGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysLeu 320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 483 GTCCCTTCATTGTTGAAAAGACACAAGAGTCTGCGCAAGAGATACCATGGAAGT 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 321 GluLysLysMetArgGluSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
    ::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 CAGAGAAGATCAGAGTGGGAATTTGACTTCATCGAATGATCCAGGATGAAC 602
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 341 VallysAsnLeuGlyProLeuAspLysPheLeuMetLysMetLysMetLysMet 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 603 GTTGGGAAAGGGATCCATGAGCCGCTGTATCGGAATGATCCAGGATGAAC 656
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
BF940797/c
LOCUS
DEFINITION
IMAGE:3280811_3' similar to SW:SR54_BACSU_P37105 SIGNAL RECOGNITION
PARTICLE PROTEIN ?; mRNA sequence.
ACCESSION
BF940797
VERSION
BF940797.1 GI:12358117
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 711)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Dr. James Lupski
cDNA Library Preparation: Lupski Laboratory
cDNA Library Arranged by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 433.
FEATURES
Source
Location/Qualifiers
1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3280811"
/clone_lib="Lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/lab_host="DE10B"
/dev_stage="adult, 36 yr"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGCTCCG-3' and
5'-GACTAGTTCTAGATCGGAGCGGCCCTT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 193 a 145 c 88 g 283 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 1,57e-59 Length: 711
Score: 630.00 Matches: 134
Percent Similarity: 77.45% Conservative: 48
Best Local Similarity: 57.02% Mismatches: 45
Query Match: 27.70% Indels: 8

```

```

Db: 12 Gaps: 3
US-09-943-108A-2 (1-455) x BF940797 (1-711)
Qy 202 LeuMetAsnGluLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeuVal 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 708 TTAATGCCAGAGGTAAAGATGTAAGAAAGTTTCAAAATCCTAGTGAATATTATTAGTA 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 222 ValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeu 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 GTTGATGCTATGACGCCACCCAGATCTGTTAATATAGCAGAGACATTAATAATCAATCA 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 242 AspValThr-GlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAlaLe 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 GATCTTCAGNGAATATATTACAAAGTTAGATGGTGATACAAAGGTGGTGGTCTCT 529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 261 uSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerSerGluLysLeuAs 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 TTCAATAAGAGACATTACTGGCAAGCCCAATTAATAATTTGTTGGTGAACAGATGAG 469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 281 pGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspVa 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 TCATATTGAGTATTCATCCAGATAGATGGCTTCAAGAAATATTAGGAATGGGAGATGT 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 lLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGl 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 CTTTCTTTAATAGAAAAGCTCAGCAAGCTATTGACCAAGATGACAGCAASTAAGTTAAG 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 321 uLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnVa 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 TGAATAAATGTTAAATCAAGAATTTAACTTTGATGACTACTTATCAGCTATGATCAAT 289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 341 lLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsn-Lys---M 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 GAAAAGCTTGGACCTATAATAATGATAGAGATGATCCAGGTGTTAAACAAGGAAC 229
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 360 eLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleI 380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 228 TTGAGGCAATTGAT---TTTCTCAAGGAGAAACAAATGGCTACAGTTAAAGCAATCA 172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 380 leGlnSerMetThrProAlaGluArgAsnProAspThrLeu-----AsnValSerA 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 171 TACAATCGATGACAGCTAAAAGAAAGAAACAACTTCCTTAGTAATAGGAATGTTCTCA 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 398 rGlyLysArgIleAlaLysGlySerGlyArgSerLeuGlnGlu-ValAsnArgLeuMet 417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 111 GAAAGAGAAGAAATAGCTAAAAGGTTCTGGTACACAGTACAAAGAGNATAATAAGTTCTT 52
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 418 LysGln-PheAsnAspMetLysLysMetMetLysGln 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51 AAAAGGCTATGAATGATGAAAAAGCAATGAAGCAA 15
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 6
AK011928
LOCUS
DEFINITION
Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610209C12:signal recognition particle 54
kDa full insert sequence.
ACCESSION
AK011928
VERSION
AK011928.1 GI:12848353
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:2610209C12.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
AUTHORS
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636

```



2	REFERENCE	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
	AUTHORS	Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
	TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
	JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	
	MEDLINE	20499374	
	PUBMED	11042159	
3	REFERENCE	Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.	
	AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.	
	TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer	
	JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)	
	MEDLINE	20530913	
	PUBMED	11076861	
4	REFERENCE	Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.	
	AUTHORS	Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.	
	TITLE	Functional annotation of a full-length mouse cDNA collection	
	JOURNAL	Nature 409 (6821), 685-690 (2001)	
	MEDLINE	21085660	
	PUBMED	11217851	
5	REFERENCE	(bases 1 to 2387)	
	AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Onno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.	
	TITLE	Direct Submission	
	JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohana Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)	
	COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details.	
	FEATURES	Location/Qualifiers	
	source	1..2387	
		/organism="Mus musculus"	

/strain="C57BL/6J"	
/db_xref="FANTOM DB:2610209C12"	
/db_xref="MGI:1905318"	
/db_xref="taxon:10090"	
/clone_lib="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="10 days embryo"	
1..2387	
/gene="Srp54"	
238..1752	
/gene="Srp54"	
/note="data source:MGI, source key:MGI:1346087, evidence:ISS	
putative	
signal recognition particle 54 kDa"	
/codon_start=1	
/protein_id="BAB27921.1"	
/db_xref="GI:12848354"	
/db_xref="MGI:1346087"	
/translation="MVLADLRKTKTSALRSLSNATINEEVLNMLKEVCTALLEADY NIKLVKOLRENKVSADLEEMASGLNKRKMKIQHAFKELVLDVPGVKAWPTKGRKN VINEVGLGGSGKTTCSKIAYYORQKWKTCICADTERFAGADOLKONAKFAPRY GSYTEMDPVITASEGVEKFKENFEIIVDTSGHKOEDSFEEMLOYSNAIOPDNIV YMDASIQACRAQAKFKVDVSVIVYKLDGCHAKGCGALSVAAYKSPFIPTG EHHDDPEFTQPFISKLLMGDLEGLIDKVKLEKLDNEALIEKLEKHGQFTKDMTE QPNIMKMPFQIIGMTIPGTFMSKNEQSMARLKMTIMDSNNDQELDSTDG AKVNSQPKQRIQVARGSVTRDVPQLLTQYTKFAQVYKMGGIKGLFGGDMKNV SOSOMAKLNQMAKMDPRVLHMGMGAGLOSMRQFQOAGNKMGMGPNM"	
BASE COUNT	741 a 420 c 560 g 666 t
ORIGIN	
Alignment Scores:	
Pred. No.:	2,29e-57
Score:	616.50
Percent Similarity:	52.12%
Best Local Similarity:	31.78%
Query Match:	27.11%
DB:	11
Gaps:	8
Matches:	150
Conservative:	96
Mismatches:	201
Indices:	25
US-09-943-108a-2 (1-455) x AK011928 (1-2387)	
Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20	
Db 238 ATGGTTCTTAGCAGATCTTGGAGAAAAATAACATCAGCATTTACGCTCATTTGACCAATGCC 297	
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40	
Db 298 ACCATTATCAATGAAGAGGATTAATAATGCTATGCTGAAGAAGATATGTACAGCATTTATG 357	
Qy 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60	
Db 358 GAAGCAGATGTTAATATTAACCTAGTGAACCAACTCAGAGAAAAATGTTAAGCTGCAATT 417	
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80	
Db 418 GATCTTGAAGAGATGGCATCTGGACTCAACAAAAAGAAAATGATCAGCATGCTGATTT 477	
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100	
Db 478 AAGAACTTGTAAAGCTTGTAGACCCCTGGAGTTAAGCGTGAACACCGCACTAAGGGAAG 537	
Qy 101 ProThrValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120	
Db 538 CAAAATGTCATCATCTTTGTTGGATTGCAAGGAGTGGTAAACAGCAACATGTTCAAG 597	
Qy 121 LeuAlaLeuMetMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140	
Db 598 TTACGATATTATTACCAAGGAAA---GTTTGAAGACCTGTTTATATGTGCAGATACA 654	
Qy 141 TyrArgProAlaAlaAlaAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160	
Db 655 TTCAGACGAGGAGCGCTTTGACCACTAAACAGAAATGCCAACCAAGCAAGATTCGTTTC 714	



```

|||||:  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 GCAAGCAAAATAGCACGATGTTATCATATAGATACCGCAAGGCGCTTGGCTATCGA 418

QY 199 pCUAlaLeuMetAsnGluLeuLysGluValLysGluLeuAlaLysProAsnGluLem 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 TGAAGATGATGAACGAATTTAAACGTTCCACCATGCTGTAAACACTACAGAACTCT 478

QY 219 eLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspA 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 TTTTCGTAAGTACTATGACTGGGAGGATGCTGTGAATCCAGCAAAAGCTTTCAAG 538

QY 239 sGlnLeuAsp-ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArg-GlyGI 238.
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 ATGTCCTTAATATACGGGTCTAGTCTTACTAAATTAGATGCTGATCTCGGAGGTGG 598

QY 258 yAlaAlaLeuSerLeuArgSer-ValThrGlnLysProLysPheVal-GlyMetSer 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 GCCCGCATACATCCGTCGGGTAGAGAAACCAATTAAGTTAATTTTCACCGGT 658

QY 278 GluLysLeuAspGlyLeuGluLeuPhe 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 GAGAAATGGATGCTCTGGGATATTT 685

RESULT 8
BE432304
LOCUS BE432304
DEFINITION EST398833 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone CLEG7M2, mRNA sequence.
ACCESSION BE432304
VERSION BE432304.1 GI:9430147
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 605)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Niernan,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
Contact: CUGI
Clemson University
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source 1..605
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEG7M2"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCUadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 190 a 97 c 159 g 159 t
ORIGIN

Alignment Scores:
Pred. No.: 3 86e-52 Length: 605
Score: 562.50 Matches: 105
Percent Similarity: 75.25% Conservative: 47
Best Local Similarity: 51.98% Mismatches: 49

```

```

Query Match: 24.74% Indels: 1
DB: 10 Gaps: 1
US-09-943-108a-2 (1-455) x BE432304 (1-605)

QY 65 ValMetGlnSerLeuThrProGlyGlnGlnValLysLeuValGlnAspGluLeuThr 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 TTGATTCGAGGTGTAACCAACAGATAGTAAAGACTGTAGCTGACGAGCTTG 61

QY 85 LysLeuMetGlyGlyGlnSerThrSerLeuMetSerAsnLysProThrValVal 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 AAACGTGATGGTGGAGAGTTCTTGAACCTGTTTTCGTAATCTAAGCCACCGCTAATA 121

QY 105 MetMetValGlyLeuGlnGlyLysThrThrAlaGlyLysLeuAlaLeuLeu 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 CTATGCGCGGTCTACAGGTGTTGGGAGACACTGTAGTCGAAGTAGCTTATAT 181

QY 125 MetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspLysPheArgProAla 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 CTA---AAGAAGCAGGTAAGAGTTGCTGCTGATTCCTGGAGACGTGTACAGACCTGCT 238

QY 145 AlaIleAsnGlnLeuGlnThrValGlyLysGlnLysLeuAspLysProValTyrSerGluGly 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 GCTATGTACCAACTTGTATTGTTGGTAAACAGGTTGATGCTGTTTATGCGACGAGA 298

QY 165 AspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHis 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 ACAGATGTAAACCTCGCAAAATAGCCCGCAAGATTACAGAGCCCAAAAGAGAGAT 358

QY 185 LeuAspPheValIleLeuAspThrAlaGlyArgLeuHisLysLeuAlaLeuMetAsn 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 GTAGATGTAGTCAATTAATGGATACAGCTGGAAGACTTCAGATAGATAAACTATGATGAT 418

QY 205 GluLeuLysGlnValLysGluLeuAlaLysProAsnGluLeuMetLeuValValAspSer 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 GAATTAAGACCTGGAACGGGTACTGAACCCACAGAGGTTTACTTCTGTTGATGCA 478

QY 225 MetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGlnLeuAspValThr 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 ATGACTGCCCAAGAGCTGCAGCTTGGTCACACATTCATCTCGAATTTGGAATTAAT 538

QY 245 GlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerLeuArg 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 GTGTCATCTTTCGAGAGGTAGATGGAGATTCTAGGGGTGGAGCAGCTTTAAGTGTCAA 598

QY 265 SerVal 266
|||||
Db 599 GAGGTA 604

RESULT 9
AW648731
LOCUS AW648731
DEFINITION EST327101 tomato germinating seedlings, TAMU Lycopersicon
esculentum cDNA clone cLEI5G22 5', mRNA sequence.
ACCESSION AW648731
VERSION AW648731.1 GI:7409885
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 593)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Niernan,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University

```

100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

## FEATURES

Location/Qualifiers

1. .593

/organism="Lycopersicon esculentum"

/cultivar="NA496"

/db\_xref="taxon:4081"

/clone="G1E15G22"

/tissue\_type="tomato germinating seedlings, TAMG"

/dev\_stage="7 days post imbibition"

/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:

hnoI; 7 days post imbibition on water-agar. Mixed stage

whole germinating seedlings from seed coat emergence up

to two centimeters in seeds not showing obvious signs of

germination were discarded."

191 a 97 c 159 g 146 t

## BASE COUNT

ORIGIN

## Alignment Scores:

Pred. No.: 1.19e-51 Length: 593  
Score: 558.00 Matches: 104  
Percent Similarity: 76.14% Conservative: 46  
Best Local Similarity: 52.79% Mismatches: 47  
Query Match: 24.54% Indels: 0  
DB: 10 Gaps: 0

US-09-943-108a-2 (1-455) x AW648731 (1-593)

QY 148 GlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGlyAspGlnVal 167

Db 2 CRACCTGTTATTTGGGTAAACAGGTGATGACCTGTTATGACGACGAGAACAGATGTA 61

QY 168 LysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAspPhe 187

Db 62 AAACCTGCAGAAATAGCCGACAGAGGTACAGAGGCGCAAAAGAAAGATGTAGATGA 121

QY 188 ValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLys 207

Db 122 GTCATATGATGATACAGCTGGAGACTTCAGATAGATAAACTATGATGATGATGATGAT 181

QY 208 GluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSerMetThrGly 227

Db 182 GACGTGAACGGGTACTGAACCCACAGAGGTTTACTGTTGTTGATGATGATGATGAT 241

QY 228 GlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyValThr 247

Db 242 CAAGAAGCTGCAGCTTGGTCACACATTCATCAATCTGAAATTTGGAATTAAGTGGCATC 301

QY 248 LeuThrLysLeuAspGlyAspThrArgGlyAlaAlaLeuSerIleArgSerValThr 267

Db 302 TTGACGAGCTAGATGGAGATCTAGGGGTGGAGACCTTAAGTTGTCAGAGGATGATCA 361

QY 268 GlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHis 287

Db 362 AGAAGCCATCACTCGTAGGAGGGGTGACCTGATGAGGACCTTGAACCTTCTAT 421

QY 288 ProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIleGluLys 307

Db 422 CCTGACCGCATGCTGGACGTATTTAGGAATGGGAGATGTTCTATGCTTTGTTGAGAA 481

QY 308 AlaGlnGluAspValAspGlnGluLysAlaLysAspLeuLysLysMetArgGluSer 327

Db 482 GCCAAGAAGTATTCGTCAGAGAAGATGCTGAAGATCTCGAGAAGAAGATCAAGTGCA 541

QY 328 SerPheThrLeuAspPheLeuGluGlnLeuAspGlnValLysAsnLeu 344

Db 542 AAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 592

RESULT 10

AI977909

LOCUS

581 bp mRNA linear EST 27-AUG-1999

## DEFINITION

496035E06.xl 496 - stressed shoot cDNA library from Wang/Bohnert  
lab Zea mays cDNA, mRNA sequence.

## ACCESSION

AI977909

## VERSION

AI977909.1

## KEYWORDS

EST.

## SOURCE

Ze mays.

## ORGANISM

Ze mays.

## REFERENCE

1 (bases 1 to 581)

## AUTHORS

Walbot,V.

## TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford

## JOURNAL

University

## COMMENT

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 496035 row: E column: 06.

## FEATURES

source

1. .581

/organism="Zea mays"

/cultivar="B73"

/db\_xref="taxon:4577"

/clone\_lib="496 - stressed shoot cDNA library from

Wang/Bohnert lab"

/tissue\_type="seedling"

/dev\_stage="salt stress"

/lab\_host="E.coli XL Gold"

/note="Organ: shoot; Vector: pBluescriptII SK(+)"

176 a 93 c 161 g 150 t 1 others

## BASE COUNT

ORIGIN

## Alignment Scores:

Pred. No.: 2.19e-51 Length: 581

Score: 555.50 Matches: 105

Percent Similarity: 74.61% Conservative: 39

Best Local Similarity: 54.40% Mismatches: 48

Query Match: 24.43% Indels: 1

DB: 9 Gaps: 1

US-09-943-108a-2 (1-455) x AI977909 (1-581)

QY 38 AlaLeuPheGluAlaAspValAsnPheLysValValLysGluPheLysThrValSer 57

Db 4 GCACCTTTGGAGCGCATGTAAGTTGCCAGTAGTAAGAAGATTATTGAGTCTGTAAGT 63

QY 58 GluArgAlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLys 77

Db 64 GAAAGGCTGTAGGACCGCATGTGATCCGAGGTGCCAGCTGCCAGCAGTTGGTGAAG 123

QY 78 IleValGlnAspGluThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSer 97

Db 124 GTTGTGAGTGTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 183

QY 98 AsnLysProProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThr 117

Db 184 AAATCTGGCCCAACTATTATTATTATTTGGCAGGCTGCAAGGTGTTGGAATAACTACTGT 243

QY 118 AlaGlyLysLeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAla 137

Db 244 TGTGCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300

QY 138 AlaAspIleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAsp 157

Db 301 GCAGATGTTTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360

QY 158 IleProValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeu 177

```

102 361 GTACCAAGTTTACTCAGAGAACTGAAGCAAAACCTTCACAAATAGCAAAACGGTTTG 420
103 178 LysHisAlaLysGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHis 197
104 421 AAGAGGCAAAAGTCAATAGGCTGATGTAATATAGTGGACAGCGCTGGAAGACTGCAG 480
105 198 IleAspGluAlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlu 217
106 481 GTAGTAAGCAATCATGATGATGAGTGAAGAGATGAAGAAAGACAGTGAATCCTACAGA 540
107 218 IleMetLeuValLysSerMetThrGlyGlnAspAla 230
108 541 GTTCTCTTGGTGTGATGCCATGACTGCCAGAGAGCT 579
109
110 RESULT 11
111 LOCUS BM413418 677 bp mRNA linear EST 22-JAN-2002
112 DEFINITION EST587745 tomato breaker fruit Lycopersicon esculentum cDNA clone
113 CLE6363C14 5' end, mRNA sequence.
114
115 ACCSSION BM413418
116 VERSION BM413418.1 GI:18265048
117 KEYWORDS EST.
118 SOURCE tomato.
119 ORGANISM
120 Lycopersicon esculentum
121 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
122 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
123 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
124 Lycopersicon.
125 1 (bases 1 to 677)
126 AUTHORS Alcaldia, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai
127 J., Bougri, O., Kirkness, E., Utterback, T., Van Aken, S., Renning
128 C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
129 Generation of ESTs from tomato fruit tissue, breaker stage (2002)
130 Unpublished (2002)
131 COMMENT Contact: CUGI
132 Clemson University Genomics Institute
133 100 Jordan Hall, Clemson, SC 29634, USA
134 Email: http://www.genome.clemson.edu/orders/index.html
135 This clone is available through the Clemson University Genomics
136 Institute
137 Seq primer: T3.
138
139 FEATURES
140 source Location/Qualifiers
141 1..677
142 /organism="Lycopersicon esculentum"
143 /cultivar="TA496"
144 /db_xref="taxon:4081"
145 /clone="CLE63C14"
146 /tissue_type="Pericarp"
147 /dev_stage="breaker"
148 /lab_host="SOLR"
149 /note="Vector: pBluescriptSKmCquadapt; Site 1: EcoRI;
150 Site 2: XhoI; supplier: Boyce Thompson Institute; Fruit
151 sequencing: The Institute for Genomic Research. Fruit
152 were harvested at the breaker stage (first sign of
153 lycopene accumulation on the blossom end of fruit). Fruit
154 were cut in half and the seeds and locules were discarded
155 prior to freezing the pericarp."
156
157 BASE COUNT 222 a 119 c 165 g 171 t
158 ORIGIN
159
160 Alignment Scores:
161 Pred. No.: 6 42e-49 Length: 677
162 Score: 534.00 Matches: 101
163 Percent Similarity: 75.38% Conservative: 46
164 Best Local Similarity: 51.79% Mismatches: 47
165 Query Match: 23.48% Indels: 1
166 DB: 13 Gaps: 0
167
168 US-09-943-108a-2 (1-455) x BM413418 (1-677)

```

```

169 152 ValGlyLysGlnIleAspIleProValTyrSerGluGlyAspGlnValLysProGlnGln 171
170 92 GTGGGGAACAGGTTGATGCTACCTGTTATGCGAGCAGACAGATGTAACCACTGAGAA 151
171 172 IleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAspPheValIleLeuAsp 191
172 152 ATAGCCCGCAACAGGATTACAGAGAGGCAAAAGAAAGATGTAGATGTAGTCAATAATGGAT 211
173 192 ThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLysGluValLysGlu 211
174 212 ACAGCTGGAGAGCTTCAGATAGATAAATACTATGATGGTATTAAGAGACGTGAACGG 271
175 212 IleAlaLysProAsnGluIleMetLeuValLysSerMetThrGlyGlnAspAlaVal 231
176 272 GTACTGAACCCACAGAGGTTTACTGTTGTTGATGCAATGCACTGGCCAGAGAGTGCA 331
177 232 AsnValAlaGluSerPheAspAspGlnIleLeuAspValThrGlyValThrLeuThrLysLeu 251
178 332 GCTTGGTCACACAACTTCATCTCGAAATTTGGAATTTACTGGTCCCATCTTGACGAAGCTA 391
179 252 AspGlyAspThrArgGlyGlyAlaLeuSerIleArgSerValThrGlnLysProIle 271
180 392 GATGGAGATCTAGGGGTGGAGCAGCTTTAAGTGTCAAGGAGGTATCAGGAAGCAATC 451
181 272 LysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMet 291
182 452 AAGCTCGTAAGAGGGGTGAACGTATGGAGGACCTTGAACCTTCTATCTGCTGACCGCATG 511
183 292 AlaSerArgIleLeuGly-MetGlyAspValLeuSerLeuIleGluLysAlaGlnGlnAs 311
184 512 GCTGGAGCTATTATTAGGAATGGGAGATGTTCTATCGTTGTTGAGAAAGCCCAAGAG 571
185 311 pValAspGlnGluLysAlaLysAspLeuGluLysLysMetArgGluSerPheThrLe 331
186 572 TATCGCTCAGAGAGATGCTGAGAAATTGCAGAGAGAGATCATGAGTGCATAAATTGATT 631
187 331 uAspAspPheLeuGluGlnIleLeuAspGlnValLysAsnLeuGly 345
188 632 CAATGACTTCTGAAGCAAACTCCGCGAGTTGCTAGTAGTGGT 674
189
190 RESULT 12
191 AY109321
192 LOCUS Zea mays P00070230 mRNA sequence.
193 DEFINITION Zea mays P00070230 mRNA sequence.
194 ACCESSION AY109321
195 VERSION AY109321.1 GI:21212876
196 KEYWORDS HTC.
197 SOURCE Zea mays.
198 ORGANISM Zea mays.
199 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
200 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
201 clade; Panicoideae; Andropogoneae; Zea.
202 1 (bases 1 to 1200)
203 AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
204 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
205 Maize Mapping Project/DuPont Consensus Sequences for Design of
206 Overgo Probes
207 Unpublished (2002)
208 REFERENCE 2 (bases 1 to 1200)
209 COE, E.C.
210 Direct Submission
211 Submitted (25-APR-2002) Maize Mapping Project, University of
212 Missouri, Columbia, MO 65211, USA
213
214 FEATURES
215 Location/Qualifiers
216 1..1200
217 /organism="Zea mays"
218 /db_xref="taxon:633818"
219 /db_xref="maizeDB:633818"
220 /clone="pcc070230"
221 /clone_lib="Maize Mapping Project/DuPont Consensus
222 Library"
223 /note="this sequence is part of a project of EST
224 assemblies resulting from the application of public

```

contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 387 a 216 c 317 g 279 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 4.15e-47 Length: 1200  
Score: 520.50 Matches: 102  
Percent Similarity: 70.82% Conservative: 63  
Best Local Similarity: 43.78% Mismatches: 65  
Query Match: 22.89% Indels: 3  
DB: 11 Gaps: 2

US-09-943-108a-2 (1-455) x AY109331 (1-1200)

QY 208 GlnVallysglnleAlaLysProAsnGlnleMetLeuValValAspSerMetThrGly 227  
DB 11 GAAGTGAAGAAGCTGTTAACTCCACAGAAGTTCTGCTGCTGTAAGCATGACTGGC 70  
QY 228 GlnAspAlaValAsnValAlaGlnSerPheAspAspGlnLeuAspValThrGlyValThr 247  
DB 71 CAAGAAGCAGACGACACTAGTCACCACTTCAATATTGAGATTGGTATTAAGTGGTGAATA 130  
QY 248 LeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArgSerValThr 267  
DB 131 CTGACTAAATTGGATGGTGACTCAGGGGGGAGCGGAGCTAAGCGTTAAAGAGGTCTCT 190  
QY 268 GlnLysProIleLysPheValGlyMetSerGlnLysLeuAspGlyLeuGluLeuPheHis 287  
DB 191 GGGAGGCCATAAATTTGTTGGCGTGGGACGATAGAGGACCTTGACCTTTCTAC 250  
QY 288 ProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIleGluLys 307  
DB 251 CCTGATCGCATGCCACGAGGTTCTGGGAATGGGAAAGTGTGTCTATTTGTGCAAAAG 310  
QY 308 AlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeuGluLysMetArgGluSer 327  
DB 311 ACACGACATTATTGGAATGATGCCAGGCATGAACAAGATAACTCCCGCA---CAAATCCGA 370  
QY 328 SerPheThrLeuAspAspPheLeuGlnLeuAspGlnValLysAsnLeuGlyProLeu 347  
DB 371 AAATTGACTTCAACGATTCTTAAACAATCTCAAAACGTTGCGAATAATGGGTTCGATG 430  
QY 348 AspAspIleMetLysMetIleProGlyMetAsnLysMetGlyLeuAspLysLeuAsn 367  
DB 431 AGCCGCATTATTGGAATGATGCCAGGCATGAACAAGATAACTCCCGCA---CAAATCCGA 487  
QY 368 MetSerGlnLysGlnIleAspHisIleLysAlaIleIleGlnSerMetThrProAlaGlu 387  
DB 488 GAACCTGAGAAAGACTTCATCTGGAGTCAATGATCAATGCCATGACTGCTGAGGAA 547  
QY 388 ArgAsnAsnProAspThrLeuAsnValSerArgLysLysArgile-----AlaLysGly 405  
DB 548 AGGAGAAGCCAGAGTACTGGCTGAATCACGTGAGAGGAGGATAAGAGTGGCTGAGGAG 607  
QY 406 SerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPheAsnAspMetLysLys 425  
DB 608 TCTGGAAGACTCAACAAGAGGTGAGTCAATTGGTTGCCAGCTTTTCCAAATCGGTGCT 667  
QY 426 MetLysGlnPheThrGlyGlyLysLysLys 438  
DB 668 CAGATGCAAGAGTGTATGGGTATGTGCAAGGACAGAA 706

RESULT 13  
AW930742  
LOCUS AW930742 533 bp mRNA linear EST 18-MAY-2001  
DEFINITION EST356585 tomato fruit mature green, TAMU Lycopersicon esculentum  
cDNA clone cLEF40N15 5', mRNA sequence.  
ACCESSION AW930742  
VERSION AW930742.1 GI:8106143  
KEYWORDS EST.

SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.  
REFERENCE 1 (bases 1 to 533)  
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.  
TITLE Generation of ESTs from tomato fruit tissue  
JOURNAL Unpublished (1999)  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
source  
1..533  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEF40N15"  
/clone\_lib="tomato fruit mature green, TAMU"  
/tissue\_type="fruit pericarp"  
/tissue\_stage="mature green (3-5 days pre-ripening)"  
/lab\_host="SOLR"  
/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; cLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"  
BASE COUNT 171 a 86 c 144 g 131 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 8.86e-47 Length: 533  
Score: 513.50 Matches: 97  
Percent Similarity: 76.27% Conservative: 38  
Best Local Similarity: 54.80% Mismatches: 41  
Query Match: 22.58% Indels: 1  
DB: 10 Gaps: 1  
US-09-943-108a-2 (1-455) x AW930742 (1-533)  
QY 108 GlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuMetArgLys 127  
DB 5 GGTCTACAGGTGTGGGAAGACAACCTGTTAGTGCAAAGTAGCTTATATCTA--AAG 61  
QY 128 LysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArgProAlaAlaIleAsn 147  
DB 62 AAGCAGGTAAAGATGGATGCTGATTCGTGGAGACTGTACAGACCTGCTGCTATTGAC 121  
QY 148 GlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGlnLysAspGlnVal 167  
DB 122 CAACTGTTATTTGGTAAACAGGTGTGATCTACCTGTTTATGTCAGCAGGACAGATGTA 181  
QY 168 LysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAspPhe 187  
DB 182 AAACCTGCAGAAATAGCCGCAAGGATTACAAGAGGCCAAAGAGAGATGATGATGA 241  
QY 188 ValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLys 207  
DB 242 GTCATAATGATACAGCTGGAAGACTTCAGATAGATAAAACATATGATGATGAATTA 301  
QY 208 GlnValLysGlnIleAlaLysProAsnGlnIleMetLeuValValAspSerMetThrGly 227  
DB 302 GACGTGAACCGGTACTGAACCCACAGAGGTTTACTTGTGTGATGCAATGACTGGC 361  
QY 228 GlnAspAlaValAsnValAlaGlnSerPheAspAspGlnLeuAspValThrGlyValThr 247  
DB 11 Gaps: 2

```

Db 362 CAAGAAGCTCAGCTTTGGTCACAAATCAATCAATCGAAATTGGAAATTACTGTGTCATC 421
QY 248 LeuThrLysLeuAspGlyAspThrArgGlyValAlaLeuSerIleArgSerValThr 267
Db 422 TTGACGAGCTAGATCGAGATCTCAGGGGUGAGCAGCTTTAGTGTCAAGAGGATCA 481
QY 268 GlnLysProTleLysPheValMetSerGlnLysLeuAspGlyLeuGlu 284
Db 482 NGAAGCAATCAAGCTCGTAGGAGGGGTGAACGTATGAGGACCTTGAA 532

RESULT 14
BQ996389 677 bp mRNA linear EST 22-AUG-2002
LOCUS QGG12107.Y3.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
DEFINITION QGG12107, mRNA sequence.
ACCESSION BQ996389
VERSION BQ996389.1 GI:22430785
SOURCE EST.
ORGANISM Lactuca sativa.
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
1 (bases 1 to 677)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmil.ucdavis.edu]
belongs to contig QG_CA_Contig5305, see http://cgpdb.ucdavis.edu/
for details.
Plate: QGG12 row: L column: 07.
FEATURES
source
location/Qualifiers
1..677
/organism="Lactuca sativa"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGG12107"
/lab_host="E.coli"
/note="Vector: pBRCNDSFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB=QG_EFGHJ lettuce serriola
TAG_ISSUE=flowers pre-fertilized
TAG_SEQ=CGTTCAGCGG"

BASE COUNT 213 a 100 c 192 g 172 t
ORIGIN

Alignment Scores:
Pred. No.: 7.17e-46 Length: 677
Score: 506.50 Matches: 100
Percent Similarity: 71.43% Conservative: 45
Best Local Similarity: 49.26% Mismatches: 57
Query Match: 22.27% Indels: 1
DB: 14 Gaps: 1

```

```

US-09-943-108A-2 (1-455) x BQ996389 (1-677)
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetClnLysMetArgGlyLysGlyLys 22
Db 72 TTTGGTCAGCTAACCCAGTGGCCCTTGAAATCCGCCCTGGGAACAGCTTAAGGGCGAAGAGTC 131
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
Db 132 TTAACGAGAGGAGATATCTCGGAACCAATGAGGACATTAGACGAGCTCTTCTCGAAGCA 191
QY 43 AspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db 192 GATGTAGCTCTCTCTGTTAAGAAGATTCGTTCAAACTGAAGTGAACAAGCTGTCGGT 251
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGlnAspGlu 82
Db 252 GTTGGTGTGACTAGAGGAGTCAACACAGATCGCAATGTGTTAAATGTGTAAGTGATGAA 311
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProProThr 102
Db 312 CTCGTGAAGCTAATGGAGAGAGGTTTCTGACTTAACCTTTTGCAAAATCGGCCCACT 371
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 372 GTTATCTTACTGGCTGGTCTACAGAGTGTGGGAAGACAACTGTTAGTGCAAAATTAGCT 431
QY 123 LeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIleValArg 142
Db 432 TTGTACTCTC---AAGAAACAGGGAGAGTTCGATGCTGATGCTGGAGATGTAATAGA 488
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
Db 489 CCAGCTGCGATTGATCAACTGTTATTTGGTAAACAGGTGGATGTTCTCTGTGTATGCA 548
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
Db 549 GTAGGAGCTGATGTAAACACAGCACTAGACAGAGTCTTCAAGAGCTTAAAGAG 608
QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu 202
Db 609 AAGATGTAGATGCTGTATATGATGATACACAGAGAGACTTCAGATAGACAAAATATG 668
QY 203 MetAsnGlu 205
Db 669 ATGGATGAA 677

RESULT 15
BQ323434 514 bp mRNA linear EST 04-JAN-2002
LOCUS PIC1_19_C04.bl_A002 pathogen-infected compatible 1 (PIC1) Sorghum
DEFINITION bicolor cDNA, mRNA sequence.
ACCESSION BQ323434
VERSION BQ323434.1 GI:18061215
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 514)
Cordonnier-Pratt,M.-M., Gingle,A., Fang,G.C., Dean,R., Wing,R.,
Sudman,M. and Pratt,L.H.
An EST database from Sorghum: plants infected with a compatible
pathogen
Unpublished (2002)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector, and regions

```

below Phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with PolyMix or T7 sequencing primer, are presented as the reverse complement.

Seq primer: JEN REV

High quality sequence stop: 454

POLYA-No.

# FEATURES

source

Location/Qualifiers

1..514  
 /organism="Sorghum bicolor"  
 /cultivar="BRX623"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Pathogen-infected compatible 1 (PIC1)"  
 /tissue\_type="Leaves"  
 /dev\_stage="4-week-old seedlings infected with Colletotrichum graminicola"  
 /note="Vector: pBluescript II SK(-) from Lambda Zap II; Site\_1: XhoI; Site\_2: EcoRI; Four-week-old sorghum seedlings were sprayed with spore suspension prepared from 3-week-old FRM421, a sorghum isolate of the anthracnose pathogen Colletotrichum graminicola. Inoculated plants were kept in a 25 C dark growth chamber with 100% relative humidity for 24 hr, followed by 12/12 hr of light/dark cycle at 25 C with 90% relative humidity for another 24 hr. All leaves were harvested and quick frozen with liquid nitrogen and stored in a -80 C freezer. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 159 a 91 c 137 g 127 t

ORIGIN

## Alignment Scores:

Pred. No.: 1.22e-43 Length: 514  
 Score: 485.00 Matches: 90  
 Percent Similarity: 77.25% Conservative: 39  
 Best Local Similarity: 53.89% Mismatches: 38  
 Query Match: 21.33% Indels: 0  
 DB: 13 Gaps: 0

US-09-943-108a-2 (1-455) x BM323434 (1-514)

Qy 194 GlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLysGluValLysGluIleAla 213  
 Db 10 GGAAGACTGCAGATTGATAAATCAATGATGGATGAATGAAGAGTGAAGAGGCTGTT 69  
 Qy 214 LysProAsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnVal 233  
 Db 70 ARTCCTACAGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 129  
 Qy 234 AlaGluSerPheAspGluLeuAspValThrGlyValThrLeuThrLysLeuAspGly 253  
 Db 130 GTCCACCCTTCAATATTGAGATTGGTGTATCATTGCTGCTGCTGCTGCTGCTGCTGCT 189  
 Qy 254 AspThrArgGlyClyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPhe 273  
 Db 190 GACTCCAGGGGGGAGCGGCGACTAAGTGTAAAGAGGTCTCTGGGAAGCCCATCAAGTTT 249  
 Qy 274 ValGlyMetSerGluLysLeuAspGlyLeuGluPheHisProGluArgMetAlaSer 293  
 Db 250 GTTGGGGCGTGGGAGGATGGAGGACCTTGAGCTTTTCTACCTGATCGGTCGACAG 309  
 Qy 294 ArgIleLeuGlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnGlnAspValAsp 313  
 Db 310 CGAGTATTGGGAATGGGAGATGATCTCTATTGTTGAAAAGACACACAAGAGCTTATGCGG 369  
 Qy 314 GlnGluLysAlaLysAspLeuGluLysMetArgGluSerPheThrLeuAspAsp 333  
 Db 370 CARGAAGAGGCTGTAGATTACAGAAAAGATCATGAGTGCAGAAATTCGACTTCAACGAC 429  
 Qy 334 PheLeuGluGlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMet 353

Db 430 TTCTTAAAAACAATCTCAAAATGTTGGAAAATGGTTCATGAGCCGCAATTATTGGAATG 489

Qy 354 IleProGlyMetAsnLysMet 360

Db 490 ATGCCAGGCATGAACAAGATA 510

Search completed: February 25, 2003, 03:09:37  
 Job time : 1571 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 00:20:56 ; Search time 3139 Seconds  
(without alignments)  
4218.470 Million cell updates/sec

Title: US-09-943-108a-2  
Perfect score: 2274  
Sequence: 1 MAFGLSERLQATQMKRGK.....GKGRNQMNLMKGNLPF 455

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xip  
-O=/cgn2.1/USPTO.spool/US0943108/runat\_21022003\_151630\_1801/app\_query.fasta\_1.647  
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0943108.cgn\_1\_1.3745.0runat\_21022003\_151630\_1801 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.inu.\*  
19: em.ju.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rod.\*  
36: em.htg\_nam.\*  
37: em.htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2274	100.0	1368	6	E36050 Signal reco
2	2274	100.0	303750	1	AP003133
3	2274	100.0	347235	1	AP003361
4	2261	99.4	295350	1	AP004826
C 5	1719.5	75.6	2256	6	AX416333
C 6	1717.5	75.5	347050	1	AL591981
C 7	1708.5	75.1	313450	1	AL596170
8	1708.5	75.1	319630	6	AX413016
C 9	1708.5	75.1	349980	6	AX417046
10	1647	72.4	2739	1	D14356
11	1647	72.4	208780	1	BSU80009
C 12	1584	69.7	303249	1	AP001515
13	1402	61.7	1136	6	AX433432
14	1393	61.3	11024	1	AE007684
C 15	1382	60.8	296750	1	AP003191
C 16	1352.5	59.5	14571	1	AE013104
C 17	1315	57.8	11023	1	AE006391
18	1312	57.7	1605	6	AX194269
C 19	1307	57.5	10462	1	AE008489
C 20	1306	57.4	1572	6	AX194043
C 21	1306	57.4	7577	6	BD003726
C 22	1306	57.4	9985	1	AE007428
23	1305	57.4	792	6	E36051
C 24	1299	57.1	77743	2	SPNEU1910
25	1297	57.0	1569	6	AR081886
26	1297	57.0	1569	6	AR195303
27	1297	57.0	1569	6	B35383
28	1286	56.6	4152	1	U88582
C 29	1285	56.5	50946	1	AE014153
C 30	1283	56.4	10173	1	AE010040
C 31	1279.5	56.3	12187	1	AE006560
C 32	1238	54.4	11825	1	AE010644
C 33	1180.5	51.9	15721	1	AE001802
C 34	1173.5	51.6	21387	1	AE008821
C 35	1173.5	51.6	274050	1	AL627276
36	1168	51.4	1841	6	AX122339
C 37	1168	51.4	337200	1	AP005280
C 38	1168	51.4	349980	6	AX127150
39	1164.5	51.2	4586	1	ECRMD
C 40	1164.5	51.2	10267	1	AE005491
C 41	1164.5	51.2	11272	1	AE000347
C 42	1164.5	51.2	270365	1	AP002562
43	1150.5	50.6	10566	1	AE004142
C 44	1150	50.6	11506	1	AE004793
45	1139.5	50.1	11004	1	AE013692

# ALIGNMENTS

RESULT 1

```

E36050      E36050      1368 bp      DNA      linear      PAT 18-JUN-2001
LOCUS      Signal recognition particle polypeptide and polynucleotide.
DEFINITION
ACCESSION   E36050
VERSION     E36050.1  GI:13022452
KEYWORDS    JP 1999235183-A/1.
SOURCE      unidentified.
ORGANISM    unclassified.
REFERENCE   1 (bases 1 to 1368)
AUTHORS    Michael T.B.
TITLE       Signal recognition particle polypeptide and polynucleotide
JOURNAL     Patent: JP 1999235183-A 1 31-AUG-1999;
            SMITHKLINE BEECHAM CORP
COMMENT     OS Unidentified
            PN JP 1999235183-A/1
            PD 31-AUG-1999
            PF 03-SEP-1998 JP 1998289963
            PR 03-SEP-1997 US 60/057890, 05-MAR-1998 US 09/035382 PI
            MICHAEL TERENCE BLACK
            PC C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
            A61K31/00,
            PC A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K35/76,
            A61K39/085,
            PC A61K39/395,A61K39/395,A61K48/00,C07K14/31,C07K16/12,C12N1/15,
            C12N1/19,
            PC C12N1/21,C12N5/10,C12N5/10,C12P21/02,C12Q1/68,G01N33/15, PC
            G01N33/566//
            PC C12P21/08,G01N33/577,(C12N15/09,C12R1:445),C12N15/00,C12N5/00,
            C12N5/00,C12R1:445)
            PC (C12N15/00,C12R1:445)
            CC Strandedness: Single;
            CC Topology: Linear;
            FH Key
            FT source 1. .J368
            FT Location/Qualifiers
            FT /organism='Unidentified'.

FEATURES
            source
            1. .J368
            /organism="unidentified"
            /db_xref="taxon:32644"

BASE COUNT  521 a 186 c 288 g 373 t

ORIGIN
Alignment Scores:
Pred. No.:      6.44e-130      Length:      1368
Score:          2274.00      Matches:      455
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             6      Gaps:      0

US-09-943-108a-2 (1-455) x E36050 (1-1368)

QY 1 MetalAPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCATTGAAGCTTTATCAGACCGCTTCAAGCGACGATGCAAAATGCGTGGTAAG 60

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAACTTACTGAAGCTGATATAAGATAATATGATGCGTGAAGTAGATTACGGTTATT 120

QY 41 GluAlaAspValAsnPhelLysValLysGluPheLleLysThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAAGCGCA 180

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLleLysIleValGln 80
Db 181 TTAGGTTCCGATGTAATGCAATCATTAACACCGGCAACAAGTATTATTAATAGTTCAA 240

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAAGTGTGATGGGTGGAGAAATATACATCGATTATATGCAATAATACCA 300

```

RESULT 2

```

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrThralaGlyLys 120
Db 301 CCTACTGTGTGTATGATGGTTTCAAGGTCTGCTGTTAAACAACAACACTGCAGTAAA 360

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTATGTCGTGTAATAAATACAAACAAAACCTATGTTAGTTCGACGATATT 420

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGGATAAATCAATTCAACACAGTAGGAAACAATAATTGATATTCTGTGA 480

QY 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAAGACACACAAATTTGTAACATTAATGATTAATAACATGCT 540

QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACATTTAGACTTTTGTAAATCAATTGATACAGCAGGTCGATTACACATCGATGAA 600

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleXMetLeu 220
Db 601 GCATTTGATGAACGAATTAAGAAGACTAAAGAATTCCTTAACCAACCAAAATTAATGTTA 660

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGCAATTCATAGCAGGCTCAAGATGCTGCAATGTTGCAGATCTTTTTCACGATCAA 720

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTGTGATGTCACAGGTGTACCTTAACATAATAGATGTTACACGTGATACAGTGGTGGTCCAGCT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTATCTATTCGTCGTCGTCACACAAAACCAATTAATTTCTGCTATGAGTCAAAAGTTA 840

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGTTTTAGAGCTTAATCCATCTCTGACGATGTCATGAGTATTTAGGTATGGTGGTAT 900

QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAlaAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTTAAGTTTAAATGAAAAGCGCCACACAGATGUGGATCAAGAAAACCAAAAGATTAA 960

QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAGATGCGTGAGTCAATCGTTTACTTTAGATGATTTTGTAGAACACACTTCATCAG 1020

QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAATCTAGGACCACTGGATGATATATGAAAATGATTCACAGTATGATGAATAAATG 1080

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGGCTAGATTAAGCTTAATATGAGTGAAGAAACAAATGATCATATTAAACGGATTATC 1140

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACCGCGCTGAAGAAACAATCCACACATGAATGATCATCGTATCAATGATG 1200

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTAAAGGCTCTGGTCGTTCAATACAAGAAGTCAATCGTTTGTATGAACAATTT 1260

QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysLysLys 440
Db 1261 AACGATATCAAGAAATATGATGAACCAATTCATCGTGGCGGTAAAGGTAAAAAGGTAAA 1320

QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCATCAAAATGCAAAATATATGTTAAAGGTATGAATTTACCGTTT 1365

```

AP003133 303750 bp DNA linear BCT 02-JUL-2002  
 LOCUS Staphylococcus aureus subsp. aureus N315 genomic DNA, complete  
 DEFINITION genome, section 5/10.  
 ACCESSION AP003133 BA000018  
 VERSION AP003133.2 GI:14349175  
 KEYWORDS  
 SOURCE Staphylococcus aureus subsp. aureus N315 (sub\_species:aureus N315, strain:N315) DNA.  
 ORGANISM Staphylococcus aureus subsp. aureus N315  
 Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 REFERENCE  
 AUTHORS 1  
 Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsunaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hiramatsu, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.  
 TITLE Whole genome sequencing of methicillin-resistant Staphylococcus aureus  
 JOURNAL Lancet 357 (9264), 1225-1240 (2001)  
 MEDLINE 213111952  
 PUBMED 11418146  
 REFERENCE 2 (bases 1 to 303750)  
 AUTHORS Aoki, K., Oguchi, A., Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K. and Kikuchi, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-JAN-2001) Akio Oguchi, National Institute of Technology and Evaluation, Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
 (E-mail:oguchienite.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-8423, Fax:81-3-3481-8424)  
 COMMENT On Jun 12, 2001 this sequence version replaced gi:13701012.  
 FEATURES  
 source  
 1. 303750  
 /organism="Staphylococcus aureus subsp. aureus N315"  
 /strain="N315"  
 /sub\_species="aureus N315"  
 /db\_xref="taxon:158879"  
 202. 1155  
 /gene="SA1056"  
 202. 1155  
 /gene="SA1056"  
 /note="ORFID:SA1056"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB42308.1"  
 /db\_xref="GI:13701013"  
 /translation="MKKTGLGILLMLVAGCGGNGHKLSSKSESKQFTVKKES  
 EEKDPDEKVEETEKKMGIKDAPSLDKPLTEKSFNSKGIQGWKDYKELMGKY  
 ELADYRFKDKSGSSIKDVAFFKGGKIKRKYIETHDDVKQDYWYVDPDGKKIGNS  
 NIPFYAEMTKYDKGLVYASVEPGSVYIHKDDAIKYDDYSKKLKSQTLKIDHKRP  
 VPSYAQIKSGFVPLTSVSEMTGSKDTKEVLPALAYFTFSPKNYEDKSNPDPKVLN  
 LVHMDFLNASSDFGNAHFVYLSKYIKEYESNVEYASDSSLK"  
 complement(1318. .1521)  
 /gene="SA1057"  
 complement(1318. .1521)  
 /gene="SA1057"  
 /note="ORFID:SA1057"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB42309.1"  
 /db\_xref="GI:13701014"  
 /translation="MQWNKVIYILLALFLGSGFIHKFYAGKNMQGILHLIFCWTGIPH  
 ILAIIASAVITVEKPADEQGNVTL"  
 1820. 2308  
 /gene="SA1058"  
 1820. 2308  
 /gene="SA1058"  
 /note="ORFID:SA1058"

hypothetical protein, similar to polypeptide deformylase"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB42310.1"  
 /db\_xref="GI:13701015"  
 /translation="MAKKKLVPASHPIILTKAAQAVTKFDDSLKRLQDLDTMYAQEA  
 AKLCAPOINSGLOWAIDIMEMEGLLQVNPKIISQSNETITDLEGSIILPDVYGEVTR  
 SKMITVESYVNGNKVELTAHEDVARMLHIIDQMNGIIPTEADRIILTDKEVAYFI  
 ND"  
 2301. 3236  
 /gene="SA1059"  
 2301. 3236  
 /gene="SA1059"  
 /note="ORFID:SA1059"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methionyl-tRNA formyltransferase"  
 /protein\_id="BAB42311.1"  
 /db\_xref="GI:13701016"  
 /translation="MTKIIIFMGTPDFSTTVLEMLIAEHVDVIAVTPDPVGRKRVMT  
 PPVKVKAMKYDLPVYQPEKLSSELEQLLDVLDIVTAARGQLLESILALPKLG  
 AINVHASLLPYRGGAPIHQALIDGEQETGITIMYVKKLDAGNIISQAIKIENDN  
 VGTMDKLSLVGADLLKETLPSLIESTNESVPQDTQATFASNIRREDRIISNKPGR  
 QVFNQIRGLSPWPVAYTMDDTNLKIYDAELVETNKINEPFTIIEFTTKAIIAIVPNDN  
 EAVAIKQKLAGKRMLAANALSGAQNTLVGKKLI"  
 3155. 4540  
 /gene="SA1060"  
 3155. 4540  
 /gene="SA1060"  
 /note="ORFID:SA1060"  
 hypothetical protein, similar to RNA-binding Sun protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB42312.1"  
 /db\_xref="GI:13701017"  
 /translation="MSNEKENVSQLFKWCAKHTSRETYMIENVSLSAFDIQILN  
 EGAYSNLRINEVSENELNMDKALFTIETVYTKRTYLTDFLKPFVKTIKAWVRQ  
 LWMGIYQVYLDKVPNHAIINEAVEIAKERGVHNGVNGVILRTVMRSPLDPNEI  
 ADPKRMAIEYSMPKTIIDHWATHYGLBETETILQSLETTSTTVRANLFRASLDII  
 EKLQDEGYDEKBDHPLCYLHIGGQPIIHSRSFQDGSFVQDKSSMFVAHIMNVDRHD  
 HVLDASAPGGKACHRAEVLMPGQVQDASDIHDKIDILNFKKLRLINIKAFQHDA  
 TKPYDKYKILVDAPCGSLGVNMRKPEIKYQSKQHIESLVELQLEILENVKNNVKI  
 GGEIITCTCTELENENVIYTLNKNKNFEFPFQHPITGELVKTLQIMPQDFNSDG  
 FETTKIKRKN"  
 4543. 5637  
 /gene="SA1061"  
 4543. 5637  
 /gene="SA1061"  
 /note="ORFID:SA1061"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB42313.1"  
 /db\_xref="GI:13701018"  
 /translation="MITAEKKKNKFLPNFQKQSIYSRDEMQNNLVQGGQOKTRAK  
 QIFEWLYKRYDSIDENLSKDLQLLDNFTVTLTVVVKQESKDGTIKELFELQD  
 GTIETVLMRHDYNGSVCTVQVCGICTGCATGLGKRNLEAGEIVSVLYQKA  
 LDATERVSOIVIGICEPENYDNMDFLRIVNDNSLNGARHITVSTSGIIPRIY  
 DEADEDIQINFAVSLHAKDEVPRLMPINRAYNVEKLEIAIOYQOKTRNRTPEYG  
 LFGVNDQLEHARELAHLKGLNCHVNLIPIVHPYERNYKTKANDLIFKEKELKRLS  
 INATIRFQGSDDIADACQURAKEROVETR"  
 5644. 6387  
 /gene="SA1062"  
 5644. 6387  
 /gene="SA1062"  
 /note="ORFID:SA1062"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB42314.1"  
 /db\_xref="GI:13701019"  
 /translation="MLEAQFETDTGQHRDKNEDAGGIFYNQTNQQLLVLCDSMGQHKH  
 GEVASKFVTDELKSRFEAENLIEHQENLNNKIDINFQLYHAQENAEYKMGTT

gene CVCALVFSEKSVIVANQDSRAYVINSRQIEQITSDHSFVNHVLVLTGQITPEBAFTHPQ  
RNIITKVMGTDKRSYDPLFIKRLNFYDILLNSDLGTDYVKDNEIKRLLVKGGTIEDH  
GQQLMQLADNHSDNVTFLIAAIEGKV"

6384. 8378

/gene="SA1063"

CDS 6384. 8378

/gene="SA1063"

/note="ORFID:SA1063"

/codon\_start=1

/transl\_table=1

/product="protein kinase"

/protein\_id="BAB42315.1"

/db\_xref="GI:13701020"

/translation="MIGIKINERYKIYDKLGGGSMSTVLAEDTILNIKVAIKALFIP  
PREKETILKRERVSHNSOLSHONIVSMIDVDEDDCYILVMEYIEGTULGEYTESH  
TSLTDTAINTFNOLLGDKIKAHDMRIVRHDKIPONILIDSNKTLKIFDGTAKALSE  
GSLVDTNHLVLTQVYFSPQKAGEATDECTDIYSIGIVLEMLVGPFPNGSTAYSIA  
IKHIQSPVNTTDRKDIPOSLNSVILRATEKDKANKIKTIQMKDDLLSSVHENRA  
NEDVYELDKMTIAVPLKKEDLAKHISEKSNQPKRETTQVPIVPAHQHQQFQPEG  
TYPEPKPKKSTRKIVLLSLIFSLMIALVSFVAMFNGKYIETPDVIGKVKAEQ  
IPNNKLKIGISRSYSDKYPENEIITKTPNTGERVERGSDVDVTSKGPVKMPNV  
IGLPEALOKLSLGLKDYTIIEKVNNQAPKGYTANGSTANTETIAIHDSIKLYES  
LGTKOVYDEFEHSFKAKALEEPKVESKESYSDDDIDEGDVISQSPKESYDES  
SISVFWGKKKSSDVKTTSVDPVITGKNDSKQVAVIINDKNDGSGTEKGSFD  
ITSQRIDIPURIEIKBKGTASYIIVKVDGTVAEKEVSYDDI"

8306. 9481

/gene="SA1064"

CDS 8306. 9481

/gene="SA1064"

/note="ORFID:SA1064"

/codon\_start=1

/transl\_table=1

/product="conserved hypothetical protein"

/protein\_id="BAB42316.1"

/db\_xref="GI:13701021"

/translation="MKTGRIVKISGVQYQVNGRPNTRKGLFRKKKSPVGDIV  
EFEVNINEGHIQVFKKMKLRPPVSNIDFLVIVMSAVEPNFSTGLDRGLVIAHS  
YQNAKRLVTKKDTPIEKQFEINELIKIYENIGYETEFIGNDDRRKKIYEAPAGLI  
VLSGQSGVKSTFNHRYPELNLNDISKLNRKGFTHRHVELFERQNGYIADTPGF  
SALDPDHDKDEIKDYFLELNRGCTCFKNCNHIKPCNVCNKHOLEIGNIAQFYDH  
YLQFPNETSNKRY"

9482. 10126

/gene="cfxE"

CDS 9482. 10126

/gene="cfxE"

/note="ORFID:SA1065"

ribulose-5-phosphate 3-epimerase homolog"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB42317.1"

/db\_xref="GI:13701022"

/translation="MTKLIPSLISVDFDLQHELKRLEAGVDGVHFDYMDGQFVPMI  
STGLPILDAVRKGTLPIDVHLMENPEKYIASFAEHGADMISLHVESPHIHRALQM  
IKHLKKAGVINGTPIQSIEFILDIIVLYVMVNFPGGQSFIDOCVEKTAGLNA  
IKMEROLFIEVDGNTDTAKVCVENGATMLVTGSEFFFKQEDYKKVQTQALRG"

Alignment Scores:

Pred. No.: 3 07e-127 Length: 303750  
Score: 2274.00 Matches: 455  
Percent Similarity: 100.00 Conservative: 0  
Best Local Similarity: 100.00 Mismatches: 0  
Query Match: 100.00 Indels: 0  
DB: 1 Gaps: 0

US-09-943-108a-2 (1-455) x AP003133 (1-303750)

Qy 1 MetaLapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20

|||||

Db 26306 ATGGCATTTGAAGCTTATCAGAACGCTTGCAGACGACGATGCAAAAAATGGTGGTAAG 26365

|||||

Qy 21 GlyLysLeuThrGluAlaaspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

|||||

Db 26366 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGGAAGTAAGATTATTC 26425

Qy 41 GluAlaaspValasnPheLysValLysGluPheIleLysThrValSerGluArgala 60  
|||||

Db 26426 SAGSGTCAGCTAACTTTAAAGTGGTAAACAAATTTATTAACACAGTATCAGACGGCA 26485

|||||

Qy 61 LeuGlySeraspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGln 80

|||||

Db 26486 TTAGTTCGCGATGATGAATCAATCAACACGCGCAACAAGTTATTAATAATAGTTCAA 26545

|||||

Qy 81 AspGluLeuThrLysLeuMetGlyGluasnThrSerIleasnMetSerAsnLysPro 100

|||||

Db 26546 GATGAATTAACGAAGTTGATGGGTGGAGAAATATACATCGATTAAATGTCATAAATAACCA 26605

|||||

Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120

|||||

Db 26606 CCTACTGTGTATGATGTTTCAAGGTCTGGTAAACACAACTATGTAGTTGCAGCATATT 26665

|||||

Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaaspIle 140

|||||

Db 26666 TTAGCATTTATTGATGGTAAAAAATACACAAAAAACCCTATGTAGTTGCAGCATATT 26725

|||||

Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleaspIleProVal 160

|||||

Db 26726 TATGTCGACGACGATAAATCAATACACAGTAGGGAACAAATTCATATTCCTGTA 26785

|||||

Qy 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisala 180

|||||

Db 26786 TACAGTGAAGGAGAGTCAAGTAAGCCACACAAATTTGAACATTAATCAATAAATGCT 26845

|||||

Qy 181 LysGluGluHisLeuaspPheValIleIleaspThrAlaGlyArgLeuHisIleaspGlu 200

|||||

Db 26846 AAAGAAGAACAATTTAGACTTTGATCATCATGATACACAGGTCGATTACACATCATGAA 26905

|||||

Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220

|||||

Db 26906 GCATTGATGAACAGATTAAGAGTAAGAAATAATGCTAAACCAACCAATTAATGTTA 26965

|||||

Qy 221 ValValaspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspaspGln 240

|||||

Db 26966 GTTGTGCGATTCAATGACGGGTCAAGATGCTGCTCAATGTTGCAGATCTTTTGCAGTCAA 27025

|||||

Qy 241 LeuaspValThrGlyValThrLeuThrLysLeuaspGlyaspThrArgGlyGlyAla 260

|||||

Db 27026 CTTGATGTACAGGTGTACCTTAATAATAGATGCTGATACAGTGTGTCGTCGACGT 27085

|||||

Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280

|||||

Db 27086 TTATCTATTCGTCGGTGACACAAACCAATTAATTTGTTGGTATGAGTGAAGAAGTTA 27145

|||||

Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyasp 300

|||||

Db 27146 GATGTTTTAGAGCTATTCCATCCTGAACGATGTCATCACGATTTTAGTATGGTGCAT 27205

|||||

Qy 301 ValLeuSerLeuIleGluLysAlaGlnAspValaspGlnGluLysLysaspLeu 320

|||||

Db 27206 GTCTTAAGTTTTAATTTGAAAAGCGCAACAGATGTGATCAGAAAAGCAAAAGATTTA 27265

|||||

Qy 321 GluLysLysMetArgGluSerSerPheThrLeuaspPheLeuGluGlnLeuaspGln 340

|||||

Db 27266 GAGAAAAAGATGCGTGAGTCACTGTTTACTTTAGATGATTTTTTGAACACATTCATCAG 27325

|||||

Qy 341 ValLysAsnLeuGlyProLeuaspPheMetLysMetIleProGlyMetAsnLysMet 360

|||||

Db 27326 GTGAAAAATCTAGGACCCACTGGATGATATTGAAATGATTTCCAGGTATGAATAAATG 27385

|||||

Qy 361 LysGlyLeuaspLysLeuasnMetSerGluLysGlnIleaspHisIleLysAlaIle 380

|||||

Db 27386 AAAGGCTTAGATAAGCTTAATATGAGTGAAGCAAAATTCATCATTTAAACGATTATC 27445

|||||

Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProaspThrLeuasnValSerArgLysLys 400

|||||

Db 27446 CAGTCAATGACGCGGCTGAAGAAACCAATCCAGACACATTGAATGTATCAGTAAAAAG 27505

```

QY 401 ArgilealalysGlySerClvArgSerleuGlnGluValasnArcLeuMetLysGlnPhe 420
DB 27506 CGTATTCTGAAGGCTGGTGGTTCATTACAAAGAGTCAATCGTTTATGAACAATTT 27565
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
DB 27566 AACCATATGAAGAATATGATGAACAATTCACGTGGTGGGTAAAGGTAAAAAGGTAA 27625
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
DB 27636 CGCAATCAATGCAAAATATGTATTAAGGTATGAATTTACCGTTT 27670

RESULT 3
AP003361
LOCUS
DEFINITION
  Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
  sequence, section 4/9.
ACCESSION
  AP003361 BA000017
VERSION
  AP003361.2 GI:14246761
KEYWORDS
  Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
  strain:Mu50) DNA.
ORGANISM
  Staphylococcus aureus subsp. aureus Mu50
REFERENCE
  Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS
  Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
  Matsunaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,
  Mizutani-Uji, Y., Takahashi, N.K., Sawano, T., Inoue, R., Kaito, C.,
  Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,
  Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,
  Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.
  Whole genome sequencing of methicillin-resistant Staphylococcus
  aureus
JOURNAL
  Lancet 357 (9264), 1225-1240 (2001)
MEDLINE
  21311952
REFERENCE
  2 (bases 1 to 347235)
AUTHORS
  Ohta, T.
TITLE
  Direct Submission
JOURNAL
  Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
  of Medical Technology and Nursing, Department of Medical
  Technology, 1-1-1 Ten-noodai, Tsukuba, Ibaraki 305-8577, Japan
  (E-mail:tohta@tsakura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,
  Fax:81-298-53-3454)
COMMENT
  On May 29, 2001 this sequence version replaced gi:13875305.
FEATURES
  Location/Qualifiers
    source
      /organism="Staphylococcus aureus subsp. aureus Mu50"
      /strain="Mu50"
      /sub_species="aureus Mu50"
      /db_xref="taxon:158878"
    35..1021
      /gene="SAV0992"
    35..1021
      /gene="SAV0992"
      /codon_start=1
      /transl_table=11
      /product="oligopeptide ABC transporter ATP-binding protein
      homologue"
      /protein_id="BAB57154.1"
      /db_xref="GI:14246762"
      /translation="MNNVLEVKDLETSLKINNEWATVENISPELSKGVILGIVES
      GCGSKTSLKSIKLLPKSKISLSSGEVIFDGRIDTLEKQLLDIRGNDIAMIFQPM
      TALNPVTKQVLESISKSHKISKKEANNLAKDLLKKVIAQDELNSYPHQLSGG
      MRGVNANLSCSPKLLIADPTLDTIQQLDLKELQKQETQMAIMMTHDLS
      VVAEFCDKVLVMYAGQIVFEGGIEILHNPHTQKLLSTIPKLKGQRLETLEGI
      VPSIQAFHNKCRFRANRCKLDCNQQSPKMHVCEIVIVRCHLYKNXEYKEI"
    1024..2004
      /gene="appf"
    1024..2004
      /gene="appf"
      /note="SAV0993"
      /codon_start=1

```

```

/transl_table=11
/product="oligopeptide transport system ATP-binding
protein AppF homologue"
/protein_id="BAB57155.1"
/db_xref="GI:14246763"
/translation="MENILEVNOIKKYYIKTKLOKTOYVKAVDVSVKSGKGTQFG
LVSGGCGKSTLAKVIIRLEDATSGSIIVNGEDITRLOGKKLKKSPQOYQMFQDPYA
SLNPNQWVDIISPEILNKKYKLPKEIKKEVLYLLKCVGLSEDAIYKYAHFESGQKQ
RVGTARALARPUSLIVADPEFVSALDVSQSVQNLMLKDQEFKLSYLFIAHLSVVK
HISDYGVMYGHIVIASDKBIYENPKPYTKALISSIPQIDKHNNRIILKGLPFS
PSNPQGPCTHTRCPIAKDKCKENIPQLKDIGNEHQVACFYVDKVGDLNA"
1997..2959
  /gene="oppb"
1997..2959
  /gene="oppb"
  /note="SAV0994"
  /codon_start=1
  /transl_table=11
  /product="probable oligopeptide transport system permease
  protein"
  /protein_id="BAB57156.1"
  /db_xref="GI:14246764"
  /translation="MLKLKRLGLMPLLLISIVVFSALIQGDPSPDLQNGKIK
  QEAINAQREKLGINDSIPHQYIRWVNHVHGDLAGESIKYKRPVIDVIEIRPFTLLG
  AMSLITVIISFALGITSGRYSYSTVQVQVFNMLAIPSIAGVFAIFSPFELQ
  WPTFGSDVNLNKEGTFEYMSKYHTFLPALTLGLLSTAGYIQYLRNDIENSCKDY
  VLTARSKGLSNKKIYNKHILRNSLIPITFLGADIVSLGAGVITETIFSYNGIKLF
  LSVTGGQIDPMNMLALYLFSGLLGNLSIDITFGIDPFRSN"
  2971..3852
  /gene="SAV0995"
  2971..3852
  /gene="SAV0995"
  /note="similar to oligopeptide transport system permease
  protein Oppc"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAB57157.1"
  /db_xref="GI:14246765"
  /translation="MQNKSFPKIALSRFTHNKIMLSIIFLLIITIYSIIAPLIAP
  FVNGQDLNLTGEMTAQNLITDGGDRNFSRLIYAGRLSISLTITIGMLLIGIT
  VGVISGSGIVDTLMTITFEFVLPFPFLFAIVLNAALGSKIKNPYGSAILVLVII
  VLSGVIARLVGRKVLQEKENEYFLAAKSIGTPTPKILKLLPNLSIVYVQATLFL
  AGMIVVESGLSFLQGISKAIPSWGNSLSDAQEGSVTSGRKPTWMPALMITLTLSI
  NFVGSGLADAFNPGRGR"
  complement(3894..4883)
  /gene="trps"
  complement(3894..4883)
  /gene="trps"
  /note="SAV0996"
  /codon_start=1
  /transl_table=11
  /product="tryptophanyl-tRNA synthetase"
  /protein_id="BAB57158.1"
  /db_xref="GI:14246766"
  /translation="METLFSGIQPSGIPTIGNYIGALKQFVDQNDYDCYFCIVQHA
  ITMPQDLRLKQTRQALAIYLAGIDPKATLFIQSEVPAHQAGWMLTIASVGL
  EFWQYKDKAKQAVEGIPAGLLIYPLMAADIVLYNTNIVPVGDQKQHIELTRNLD
  RNSRYNDVLYKPEIRMPKVGGRVMSLQDPTPKSKSDDNKNFISLIDEPNVAKKI
  KSAVDSGDIILKFDNRNKGITNLISYAGLTDMPIKTEAKYEGEGYKFGKGLDAEI
  VKAFIVFQERKEYFSYNSDKLDDILDQGRKAHKVSRFTVKRMEKAMGLGRK"
  5177..5572
  /gene="SAV0997"
  5177..5572
  /gene="SAV0997"
  /codon_start=1
  /transl_table=11
  /product="conserved hypothetical protein"
  /protein_id="BAB57159.1"
  /db_xref="GI:14246767"
  /translation="MVTIFTPSCTSCRKAKAQLEHDIPTYERNIFSHELLIDBIKQ
  ILKMTEDGTETIIRSKTYQKLNVDISLPIQDIYSIQDNPGLLRPIILDNKRLQ
  VGYNEDEIRFLPRKVRTQLQEAQRWD"

```

gene

5943. .6662  
/gene="SAV0998"  
5943. .6662  
/gene="SAV0998"  
/note="similar to negative regulator of genetic competence  
MecA"

CDS

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB57160.1"  
/db\_xref="GI:14246768"  
/translation="MRIERYDDFTVVKLFITYSDIEARGFREDLWNRKRGEFFWSM  
MEINEDFVEEGLVLIQVHAEKGVETVTSKKNEDMMNSDDDATOFQEQVEL  
LAQTEGDELEELFEQRTKEAKGKROKSSARKNRTIIVKFNDELVDVNYAHS  
NPITFEEDLLVMVDGYVYAVHFDSDVQEVINDSYSQLERAYPTDRTVEYLYDIA  
KLMSHTNAQVRYPFETTE"  
6783. .7769  
/gene="SAV0999"  
6783. .7769  
/gene="SAV0999"  
/note="similar to transcription factor"

gene

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB57161.1"  
/db\_xref="GI:14246769"  
/translation="MLVALNEEKERVLTATLRKTOYFCPCVCGQVILKRGKLVISHF  
AKHLAEQKCPNETIKHYKSLILAIOMIOGCKVELEPELKEIKOIPDILINKVY  
IELQSPYIKQIQRTEGLKKMGYKVSMLLNDVYCHNKVNFHPSLFIPPIRKL  
HTFNLEKQIMFOIOIQLGGHKYVAEKNAKILIEFNAPCDYHAYLKSFAINQY  
IKYCKWNSVLEPTLSAMYQLQTDQEVYNYGYTFPEQIYIENHPTEWQLQVLDLX  
NGKSLVNDNLNYPFKKFIIVALESKTAIEIKLINNYLNICSRGNDVQLP"  
7817. .9625  
/gene="SAV1000"  
7817. .9625  
/gene="SAV1000"  
/codon\_start=1  
/transl\_table=11  
/product="thimet oligopeptidase homologue"

gene

/protein\_id="BAB57162.1"  
/db\_xref="GI:14246770"  
/translation="MSQLSREQRERKPYEYVWDLTITFKDDERAFEAKEVENELGK  
EQFGKHIGDSAEFLYNALAELEDTLTKLEKVVYVAHLKQDDPTNDKYPGMESRAHQ  
LIETKSSAMSTVEIILQIDEDIKIYFVNSYDKLQKFAFDLKLINERPHILDAETEK  
LITEADALSPSNVYGMFNADIVFEDAIKDGNAHPLTQGTPIKYLESDDRKLRES  
APRNVYKAGAHNTLGATLAGEVKKNVFNARTHNKATAREKALSNNHIPNVYDNLV  
KTVKYPILHRYTELKELGLDGLKVDLYTPLIKDIKEMPEYREAKKMWLKALEP  
MGEYLNVVKEGLNRRVYVFNKGRSGYSSGAHLNPNPILLNWSNITSDILYLVH  
EFGHSASFSKFKQPSNSDYTIFFVAEASTCNEALLSDYMDKHLDDKRLLLNQE  
LERFRATFRQTMFAEFHKKIHAIIEAGPELPTPRMNEYAKLNKLYFGDSVETEDI  
SKWSRIPHVMNYVYQYATGYSAQSLSHQILTEGRPAVDRYINEFLKGGSSNYP  
EILKNAGVDMTTPEIEQACEVFEQKLNAPFKLMKA"  
complement(10085. .10639)  
/gene="yjbM"  
complement(10085. .10639)  
/gene="yjbM"  
/note="SAV1001"  
/codon\_start=1  
/transl\_table=11  
/product="GTP pyrophosphokinase"

gene

/protein\_id="BAB57163.1"  
/db\_xref="GI:14246771"  
/translation="MKQLSYKVVSKTEKIYTYCYNEIIPKRDITTSMTCDICQIAG  
IDLEVFKDLOKSLTESLKIIDLHAREMEIEQAPSLVFFSEDEVHEGLKVEGLYPYH  
IYYIINELMGPIEKNLPKPLEYIQOQLYVMEELLTIYEWPEKLLINKELKKAIAQ  
QKIEKLKYPDGFWSKMPKIKSK"  
complement(10575. .10901)  
/gene="SAV1002"

gene

3.57e-127 Length: 347235  
2274.00 Matches: 455  
100.00% Conservative: 0

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0

US-09-943-108a-2 (1-455) x AP003361 (1-347235)

QY 1 MetalapheluglyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
DB 256730 ATGGCAITTTGAAGGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATCGTGTAAG 256789  
QY 21 GlyLysLeuThrGluAlaAspLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
DB 256790 GGTAAACTTACTGAAGCTCATATAAGATAATATGATCGCTGAAGTAAGCTTATTT 256849  
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
DB 256850 GAGGCTGACGTGAACATTTAAAGTGGTAAAGAAATTTATTAAACAGTATCAGACGCGCA 256909  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
DB 256910 TTAGGTTCCGATGTAATGCAATCATTAAACACGCGCAACAAGTTATTAAATAGTTCAA 256969  
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
DB 256970 GATGAATTAAGCAAGTTGATGGGTGGAGAAAATACATCGATTAAATATGCAAAATAACCA 257029  
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
DB 257030 CTTACTGTTGTATGATGTTGGTTTACAAGTCTGTTAAACACACAACTGCAGGTAAA 257089  
QY 121 LeuAlaLeuLeuMetArgLysLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
DB 257090 TTAGCATTTATTGATCGCTAAAAAATACAAACAAAAAAGCTATGTTAGTTCAGCATATT 257149  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
DB 257150 TATCTCCAGCAGCGATAAATCAATTACACAGTAGGAAACAAATGATATTCTCTGA 257209  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
DB 257210 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAATTAATGCATTAAACATGCT 257269  
QY 181 LysGluGluHisLeuAspPheValIleleAspThrAlaGlyArgLeuHisIleAspGlu 200  
DB 257270 AAAGAAGAACATTTAGACTTTGTATTCATTGATACAGCAGCTCGATTACACATCGATGAA 257329  
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLeuLeaLysProAsnGluIleMetLeu 220  
DB 257330 GCATTGATGACGAATTAAGAAGTAAGTAAGAAATTCGTAACCAACCAAGAAATTAATGTTA 257389  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
DB 257390 GTTGTCTGATTCAAAGACGGGTCAAGATGCTGTCAATGTTGCAGAAATCTTTTGACGATCA 257449  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
DB 257450 CTTGATGTCACAGGTGTTACCTTAACATAATTTAGATGCTGATACACCTGGTGGTGCACCT 257509  
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
DB 257510 TTATCTATTCTGTCGTCACACAAAAACCAATTAATTTGTTGGTATGATGATAAGTTA 257569  
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
DB 257570 GATGTTTAGAGCTATTCCATCTCTGAACTGATGATGATGATGATGATGATGATGATGAT 257629  
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
DB 257630 GTGTTAAGTTTAAATGAAAAAGCGCAACGAAGATGTGGAATCAAGAAAAAAGCAAAAGATT 257689  
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340  
DB 257690 GAGAAAAAGATGCGGTGATCATCGTTACTTTTACTTTTATGATGATTTTATGACAACTTGATCAG 257749

QY 341 VallysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
Db 257750 GTGAAAAATCTAGGACACTGGATGATATTATGAAATGATCCAGGTATGAATAAAATG 257809  
QY 361 LysGlyLeuAspLysLeuAsnMetSerGlyLysGlnPheAspHisIleLysAlaIleIle 380  
Db 257810 AAAGGCTAGATAGCTTATATGAGTGAAAGCAAAATGATCATATTAAAGCGATTATC 257869  
QY 381 GlnSerMetThrProIaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
Db 257870 CAGTCAATGACGCGCGCTGAAAGAAACAATCCAGACACATTCATGATCATCGTAAAG 257929  
QY 401 ArgIleAlaLysGlySerClvArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
Db 257930 CGTATTCTAAAGGGCTGGCTGTCATACAAAGAAAGTCAATCGTTTATGAACAATTT 257989  
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440  
Db 257990 AACGATATGAAGAAATGATGAACAATTCACGTGGCGGTAAAGGTAAAGAAAGGTAA 258049  
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
Db 258050 CGCAATCAATGAAAAATATGTTAAAGGTATGATTTACCGTTT 258094  
RESULT 4  
AP004826 295350 bp DNA linear BCT 02-JUL-2002  
LOCUS Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,  
DEFINITION strain:MW2, section 5/10.  
ACCESSION AP004826 BA000033  
VERSION AP004826.1 GI:21204263  
KEYWORDS  
SOURCE Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.  
ORGANISM Staphylococcus aureus subsp. aureus MW2  
REFERENCE 1  
AUTHORS Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,  
Nagai,Y., Iwama,N., Asano,K., Naimi,T., Kuroda,H., Cui,L.,  
Yamamoto,K. and Hiramatsu,K.  
TITLE Genome and virulence determinants of high virulence  
community-acquired MRSA  
JOURNAL Lancet 359 (9320), 1819-1827 (2002)  
MEDLINE 22040717  
PUBMED 12044378  
REFERENCE 2 (bases 1 to 295350)  
AUTHORS Aoki,K., Oguchi,A., Nagai,Y., Asano,K., Iwama,N., Baba,T.,  
Kuroda,M., Hiramatsu,K. and Kikuchi,H.  
TITLE Direct Submission  
JOURNAL Submitted (06-MAR-2002) Akio Oguchi, National Institute of  
Technology and Evaluation, Biotechnology Center; 2Chome 49-10  
Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan  
(E-mail:oguchienite.go.jp, URL:http://www.bio.nite.go.jp/,  
Tel:81-3-3481-8423, Fax:81-3-3481-8424)  
FEATURES  
source  
1. .295350  
/organism="Staphylococcus aureus subsp. aureus MW2"  
/strain="MW2"  
/db\_xref="taxon:196620"  
301. .1254  
/gene="MW1096"  
301. .1254  
/gene="MW1096"  
/note="ORFID:MW1096"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB94961.1"  
/db\_xref="GI:21204264"  
/translation="MKTLGCLLIMLVAGSGFGNHLKSSKSEFSKOPTVKKES  
EEKPDLEKVEELEKMGKIDPSLDKLDPLATKSFNSKSGIQGKDKYKELMGVY  
ELADYRFDKSGSSIKDVDAFFKKGKIKKVIETDDYVKQYDVIYDPDGKIGNS  
NIPVYASIMTKYKDGKLVYASVEPGSVYHKDAIKYDDYSLKLLSOLTKLDPKP

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

VPYSAIKSGVPLISVSFWTHGSKOTKDEVLPALAYTFSPKKNYEDKSNPDKVLN  
LVHMDFLNASSDFGNRAHFVVLVSKYIKYESNYETASDDSLK"  
complement(1417..1695)  
/gene="MW1097"  
complement(1417..1695)  
/gene="MW1097"  
/note="ORFID:MW1097"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAB94962.1"  
/db\_xref="GI:21204265"  
/translation="MPELKGVLKEEANVLISNSTGGKEMQNVKVIYLLALLGLSFG  
LHKYAGKNMQGLHLFCWTGIPHILAILASVITVFKPADEQGNVL"  
1919..2407  
/gene="MW1098"  
1919..2407  
/gene="MW1098"  
/note="ORFID:MW1098"  
hypothetical protein, similar to polypeptide deformylase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB94963.1"  
/db\_xref="GI:21204266"  
/translation="MAIKKLYPASHPILTKKAQAVIKFDDSLKRLLODLEDTVQAQEA  
AGLCAPQINOSLOVALIDMEKMGILLQVNPKIIQSNETITDLEGGITLPDVYGEVTR  
SKMIVRSYDVNGNKVELTAHEDVARMILHIIDQMNGIFFTERADRILITKEVEAFI  
ND"  
2400..3335  
/gene="MW1099"  
2400..3335  
/gene="MW1099"  
/note="ORFID:MW1099"  
/codon\_start=1  
/transl\_table=11  
/product="methionyl-tRNA formyltransferase"  
/protein\_id="BAB94964.1"  
/db\_xref="GI:21204267"  
/translation="MTKIIEMGTPDFSTVLEMLIAEHQDVIAVTPQDPVRKRYMT  
PPYKVKAMKYLDPYQPEKISGSELEQLQLDVLIVTAARGLLPELSLLALPKLG  
AINVHASLPKYGAGAPLHQALIDGQETGITIMVKKLDAGNIISQAALKEKDN  
VGTMDKLLSLKELTLPESIIEGTNESVPQDDQTQATFASNTRERINNNKPGR  
QVFNQIKGSLPFWPVAYTTMDTNLKIYDAELVETKNINPEPTIETTKAILATNDN  
EAVAIKDMLAGKRLMAANYLSGAQNTLVGKKLI"  
3332..4639  
/gene="MW1100"  
3332..4639  
/gene="MW1100"  
/note="ORFID:MW1100"  
hypothetical protein, similar to RNA-binding Sun protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB94965.1"  
/db\_xref="GI:21204268"  
/translation="MIENVRSIAFDFTIQDILNEGAYSNLINENVISENELNAMDKALF  
TEIYGVTKKFTLDFYLPKPVYKIRAWRQLLWMSIYQVYLDKVPENHAIINEAVE  
IAKRGYGHNGVNVGILFMRSDLPDFNEIADPKKRAIEYSPKVIIDHWATYHG  
LEKETTLQSELETTSTTVRAMLFRASLDIIIEKLODEGYDVEKDHLPLCYLHTGQP  
LIHSRSPKDFVSIQKSSWFVAHIMNVDRHDVLDACSPGCKACHIAEVLMPGQV  
DASDIHDHKLIDILNFKKILRTNIKAFQHDATKPYDKTYDKILVDAPCSGLGVNRHK  
PEIKYQSKOHIESLVLEQLLELVNKNVKGIEIYSTCTIEQLENNVIYITFLKN  
KNFEPFPQHPITGELVKTLQIMPQDPSDGFITIKRKDN"  
4642..5736  
/gene="MW1101"  
4642..5736  
/gene="MW1101"  
/note="ORFID:MW1101"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAB94966.1"  
/db\_xref="GI:21204269"  
/translation="MITAEKKKKNKFLPNFDKQSIYSLRFDQMQLWVQEQQKFRAK

```

QIFEWLYQKRVDSIDEMTNLSKDLROLLKDNFTVTTLTVVKQESKDGTKIFLYELOD
LDTETVLRHDYCNVCVITTOVGCRTGCTCASCATGLSKLNLEAGEIYSOVLTVOKA
LQATERTVSQIVIMGICEPENYDEMDFLRIVNDRDNLNIGARHITVTSGLIIRYI
DFADIEDIQINFAVSLHAKADEVRMLPINRAINVEKLEIAQIYQEKENRRVTEYG
LFGVNDQLEHARLEHLIKGLNCHVNLIPVNHPRNYVKTAKNDIFRFEKELRLG
INATIRREGSDIDACGQLRAKERQVETR"
5743. .6486
/gene="Ww1102"
CDS
5743. 6486
/gene="Ww1102"
/translation="MLEAQFFDTGQHRKKNEDAGGIYFNRNQQLLVICDGMGSHKA
GEVASKFYDELSKFEAENLIEEHOENWLNKNIKQNFQLYHAEAKYKMGMTT
CYVALFVKTSVYIANVGDSRAYVINSRQEQITSDHSFVNLVLTQITPEBAFTPO
RNIITKVGMDTKRYSPOLETKRLNFYDYLNLNSDGLTDYVKDNEIKRLLVKSGTIEDH
GQQLMQLADNHSKDNFTFILAIEGDKV"
6483. 8477
/gene="Ww1103"
CDS
6483. 8477
/gene="Ww1103"
/translation="MIGKIINERYKIIVDKGGGMSVTYLAEDTILINIKVAILFIP
PREKETILRPEREVHNSOLSQNTSMIDVDEDDCYIVMEYIEGPTLSYEIESH
GPLVDYTAINEFTNGLDQIKHAKHMLVHRDIKPQNLILDSNKTILFEPGIAKALSE
TSUTQTNVLTQVTFSPQAKGERDECTDIYSIGIVLHMLVGPFPNGSTAVSIA
IKHIDQSPVNTVDRKDIQFSLNSVILRATEKANRYKTIQEMKDDLSVILHENRA
NEDVELDKMTIAVPLKEDLAKHISEKSNQPKRETQVPIVNGPAHQBFQKPEG
TYVEFKPKKSTRKIVLLSIFSLMIALVSFVAMFNGKYEETPDVIGKSVKAEQ
IFPNKMLKIGISSYSDKYPENELIKTPTNGVERGSDVDVVISKGPVKMPNV
IGLPEALQKLKLSGLKDVYIEKVYNNQAPKGYLANQSVTANTEIAIHDSNIKLYES
LGKQVYVEDEHKSFSKAKALEKGPVKESKEYSDDIDEGDVISQSPKSKSDEG
STISFVSQKSKSDSDVKTITTESVDVPTGKNDSQKVYIYIKNDNDGSGTEKGSFD
ITSQDRIDIPRIEIKGTASYIIVKDGKTVAEKEVSYDDV"
8705. .9580
/gene="Ww1104"
CDS
8705. 9580
/gene="Ww1104"
/translation="MKTGRIVKISGVQVYNGVERFNTKRGFLFRKKKSPVWGDIY
EFVQNGVTHQVFERKNEKLRPPVSNIDTLVIVMSAVEPNFSQLDRFLVIAHS
YQNAVRVTKKDTPEKQFEINELLKIYENIYETFEFNGDDRRKKIYEAWPAGLI
VLSGSGHTGKSTFHNHYRPELNLNDISKLNKRGKHTRHVELFERQNGYIADPGF
SALDPDHDKDEIKDYFLELNRVGETCKFRNCNHIKEPNKCNVHOLEIGNIAQFYDH
YLQLENEISNKKRY"
9581. 10225
/gene="cfxE"
CDS
9581. 10225
/gene="cfxE"
/translation="MTKLYPSLLSVDFLDLOHELKRLKEACGVGHVDFMDGQFVPI
SIGLPILDAVRKGTLPIDVHLMTENPEKXIAPFAHGADMISHVESPHIIRAQIM
IKHLDKAGVINPGTPIQSLEPILDIVYVLVMTVNPFGGGSFIDQCQVEKIAGNA
IKMERQLNFDIEVDGNGVNTIDTAKVCVENGATMLVTGSFFKQEDYKKVTOALKG"

```

```

10232. 10873
/gene="Ww1106"
CDS
10232. 10873
/gene="Ww1106"
/translation="ORFID:Ww1106"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

Alignment Scores:
Pred. No.: 1.83e-126 Length: 295350
Score: 2261.00 Matches: 452
Percent Similarity: 99.78% Conservative: 2
Best Local Similarity: 99.34% Mismatches: 1
Query Match: 99.43% Indels: 0
DB: 1 Gaps: 0

US-09-943-108a-2 (1-455) x AP004826 (1-295350)

QY 1 MetalPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 26404 ATGGCACTTTGAAGGGTATATCAGAACGCTTGCACGCGCATGCGTGAAGTAAGATTACGCTTACTT 26463
QY 21 GlyLysLeuThrGluAlaAspLysLysLysMetMetArgGluValArgLeuAlaLeuPhe 40
Db 26464 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTACGCTTACTT 26523
QY 41 GluAlaAspValAsnPhelYsValLysGluPhelLysThrValSerGluArgAla 60
Db 26524 GAGGTCAGCTAAACTTTAAAGTGTAAAGATTTTATTAACACAGTATCAAGACGGCGCA 26583
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysValGln 80
Db 26584 TTAGGTTCCGATGTATGTCAATCATTAACACAGGCGCAACAGTTATTAATAAGTTCAA 26643
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysMetSerAsnLysPro 100
Db 26644 GATGAATTAACGACAGTGTATGGTGGAGAAAATACGTCGATTATATATGTCAAAATAACCA 26703
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrThrLysLys 120
Db 26704 CCTACTGTTGTATGATGGTGTGTATACAGTGTGTGTAAACAACTCGACGTTAA 26763
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 26764 TTAGCAATTATTGATCGCTAAATAATACAAACAAACCTATGTTAGTTCACACAGATATT 26823
QY 141 TyrArgProAlaAlaLysGlnGlnLeuGlnThrValGlyLysGlnLysLysProVal 160
Db 26824 TATCGTCCAGCAGCATATAATCAATTACAAACAGTAGGGAACAAATTCATATTCCTCTA 26883
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLysValThrAsnAlaLeuLysHisAla 180
Db 26884 TACAGTGAGGAGATCAAGTAAGCCACAAATTTACTAATCATTAATAACATGCT 26943
QY 181 LysGluGluHisLeuAspPheValLysLysThrLysLysLysLysLysLysLys 200
Db 26944 AAAGAAGAACAATTATAGACTTTGTATTCATTGATACAGCAGTGCATTACACATCGATGAA 27003
QY 201 AlaLeuMetAsnGluLeuLysGluValLysLysLysLysLysLysLysLysLysLys 220
Db 27004 GCATTGTATGACGAATTAAGAAAGTAAGACATTCGTAAACCAACGAAATATGTTA 27063
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 27064 GTTGCGATTCATGACGGGTCAAGTGTCTCAATGTTCCAGAAATCTTTGACAGTCA 27123
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 27124 CTTGATGTCACAGGTGTACCTTAACATAAATTAGATGTTGATACAGTGTGTGTGAGCT 27183
QY 261 LeuSerLysArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280

```



```

Db 27184 TTATCTATTCTGTTGGTGACACAAAACCAATTAATAATTTGTTGTCAGTGAAGAAAGTTA 27243
QY
QY 281 ASPLGLYLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||
Db 27244 GATGGTTTAGAGCTATTCCATCCCTGAACGATGGCATCACGATATTTAGGTATGGGTGAT 27303
QY
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
|||||
Db 27304 GTGTTAAGTTTAAATGAAAAAGCCACACAGATGGATGATCAAGAAAAAGCAAGATTTA 27363
QY
QY 321 GLULYLYSmetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
|||||
Db 27364 GAGAAAAAGATCGGTGAGTCATCGTTTACTTTAGATGATTTTATAGAACAACTTGATCAG 27423
QY
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||
Db 27424 GTGAAAAAATCTAGGACCACTGGATGATATATGAAAAATGATCCAGGTATGAATAAATG 27483
QY
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||
Db 27484 AARGGCTAGATAAGCTTAATATGAGTGAAGCAAAATTTGATCATATATAAGCGATATC 27543
QY
QY 381 GluSerMetThrProAlaGluArgAsnAspProAspThrLeuAsnValSerArgLysLys 400
|||||
Db 27544 CAGTCATGACGCGCGCTGAAGAAGAACATCCAGACACATTTGAATGTATCAGCTAAAAAG 27603
QY
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||
Db 27604 CGTATTGCTAAAGGCTGGTGGTTCATTAAGAAGAGTCAATCGTTGATGAACAAATTT 27663
QY
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
|||||
Db 27664 AACGATATGNAGAAATGATGAACATTCACCTGGTGGCGGTAAAGGTAAAGGTAA 27723
QY
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
|||||
Db 27724 CGCAATCAAAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 27768

RESULT 5
AX416333/c
LOCUS AX416333 2256 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 3324 from Patent WO0228891.
ACCESSION AX416333
VERSION AX416333.1 GI:21448790
KEYWORDS
SOURCE Listeria monocytogenes ATCC 19115.
ORGANISM Listeria monocytogenes ATCC 19115
REFERENCE
1
AUTHORS Glaser,P. and Kunst,F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 3324 11-APR-2002;
PASTEUR Institut (FR)
FEATURES
source
1..2256
/organism="Listeria monocytogenes ATCC 19115"
/db_xref="taxon:176281"
BASE COUNT 533 a 471 c 420 g 832 t
ORIGIN

Alignment Scores:
Pred. No.: 5,36e-96 Length: 2256
Score: 1719.50 Matches: 331
Percent Similarity: 86.81% Conservative: 64
Best Local Similarity: 72.75% Mismatches: 55
Query Match: 75.62% Indels: 5
DB: 6 Gaps: 1

US-09-943-108a-2 (1-455) x AX416333 (1-2256)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||
Db 1666 ATGGCATTTGAGGACTAGCTGGAAGACTCCAGAAACAACTAGAGTCATATTGAAGCAATATC 1607

```

```

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||
Db 1606 GGAAGAGTGAAGAGCTGACGTAAAGAAATGATGCGTGAGTTCGTCTGCTCTACTT 1547
QY
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
|||||
Db 1546 GAAGCGATGTTAACTTTAAAGTCGTTAAACAATTTATTAACACAGTAAGCGAAGTCT 1487
QY
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||
Db 1486 GTGCGCGCGAGCTATGAAAGCCTAACACCGCGCCAAAGTATCAAAATCGTTCAA 1427
QY
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||
Db 1426 GAAGAACTAACAGCTTAATGGCGGAGAGAAACAAATCGGAACAGCGGCCCA 1367
QY
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||
Db 1366 CCAACGCTGATTTATGATGGTAGTTTACAAGGGGCTGTAAACAACGACTTCCGGCAAA 1307
QY
QY 121 LeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAspIle 140
|||||
Db 1306 CTTCTCTAATTTATACGCAAAATATATAACCGTAAACCACTTCTCGCAGCAGATAT 1247
QY
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||
Db 1246 TACCGACCTGCGCGCAATCAACAATTAGAACACACTTGCACGCAATTAGATATGCCGTA 1187
QY
QY 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180
|||||
Db 1186 TTTTCTCTAGGAGACCAAGTAAGCGGTAGAAATCGCAACCAAGCTATTGCTAAAGCA 1127
QY
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||
Db 1126 AAAGAAAGACATTTAGATTATGTTATGATACAGCGGTCTCTCTCATATTGACGAA 1067
QY
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
|||||
Db 1066 ACCGTAATGGATGAGTTTAAACAAGTAGAAGAAATGCAACGCCAACCGAAATCTGCTT 1007
QY
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
|||||
Db 1006 GTTGTGATTCATCCATGACGTGGCAAGATGCTGTCAACGTGGCGCAAGCTTCAACGAACA 947
QY
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyArgPheThrArgGlyAlaAla 260
|||||
Db 946 TTAGAAATTTACAGGTGTCGCTGTAAOCATAATTAGACGGGTGATACACGTGTGTGTGCG 887
QY
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlyLysLeu 280
|||||
Db 886 CTTTCGATTCGTTCCGTCACCTGGCAACCAACCAATTAATTTATCGGACACTGGTGAAGAA 827
QY
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||
Db 826 GAAGCTCAGAAACTTCCATCCAGATCGTATGGCTTCCAGAAATTTCTGGCATGGCGAT 767
QY
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
|||||
Db 766 GTGCTTTCCTGATTGAAAGAGCGCAACCGATGTAGATACAGAAAAAATGAAGCTATG 707
QY
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
|||||
Db 706 GAACAAATAATGAAGACACACAGCATGACGCTAGATGACTTCTTAGAACAGTTTACAACAA 647
QY
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||
Db 646 GTAAACAAATGGGACCACTAGATGAAGTACTTATAAATGATGCCAGGGGCAACAAATG 587
QY
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||
Db 586 AAAGGCTCGACAACTGAACGCTGATGAATAACAACTAGGTCTATATTGAAGCAATATC 527

```





```

QY 321 GluLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121156 GACAAATAAGAGACACACATGACGCTAGATGACTTCTTAAACAATTACACAA 121097

QY 341 ValLysAsnLeuGlyProLeuAspAspPheMetLysMetIleProGlyMetAsnLysMet 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121096 GTAAACAATAAGGACACACATGACGCTAGATGACTTCTTAAACAATAAGGACAA 121037

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121036 AAAGCCCTGCACACATGACGCTAGATGACTTCTTAAACAATAAGGACAA 120977

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120976 AAATCAATGACCAAAATGAAAGACACACACATGATCATCATGACAGTAGAGAAA 120917

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120916 CGGATTGACGCTAGCAGCGCTCGGTTCAAGAAATTAACCGTCTCTTAACAATTT 120857

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysLysLysGlyLys 440
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120856 GCTGAATGAAATAATGATGAGCAAGCAATGCTGCTGGCGGCAAGGCAAAAGGCAAA 120797

QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120796 -----AACCCATCGCGCAATTTCAAGATGCCATTT 120767

RESULT 7
AL596170/c 313450 bp DNA linear BCT 06-JUN-2002
LOCUS
DEFINITION Listeria innocua Clp11262 complete genome, segment 8/12.
ACCESSION AL596170
VERSION AL596170.1 GI:16414292
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A.,
Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T.,
Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P.,
Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,
Dussurget, O., Entian, K.D., Fsihi, H., Portillo, F.G., Garrido, P.,
Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J.,
Jackson, D., Jones, L.M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, P.,
Kurapat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E.,
Nedjari, H., Nordiek, G., Novella, S., de Pablos, B., Perez-Diaz, J.C.,
Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N.,
Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J. and Cossart, P.
Comparative genomics of Listeria species
Science 294 (5543), 849-852 (2001)
MEDLINE 21537279
PUBMED 11679669
REFERENCE
2 (bases 1 to 313450)
AUTHORS Glaser, P., Frangeul, L. and Rusniok, C.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Glaser P., Institut Pasteur, Genomique des
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
Cedex 15, FRANCE
COMMENT E-mail: pglaser@pasteur.fr
Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
FEATURES
Location/Qualifiers
Source 1..313450
/organism="Listeria innocua"
/strain="Clp11262"
/db_xref="taxon:1642"
complement(287..1444)
/genes="lin1790"
complement(287..1444)
/genes="lin1790"
/notes="similar to transmembrane transport proteins"

```

```

/codon_start=1
/transl_table=11
/protein_id="CAC97021.1"
/db_xref="GI:16414293"
/db_xref="SPTREMBL:Q92AX8"
/translat="MGAKGFWLTWVAISGLSQGVLLPIAIIIEKGVSGVINGF
IGATGIVLGVSPFIEAPLHYKYPILVGGVLAIFAIPIWNLVFWLRL
IGDHMLHFSOTWIGAMDSKGRNMAIYGLFFSVGAIPQVLNLAKNLPI
FISGLIYLIANSLVFIRNDVGEKAVIRKISFVGSIKRSGVFKMAWVAMPPL
IETGLNATFFVGVGRDGLTMMIASSFSVGTIIFQVPIGLISDKGRDKMLP
TGAGAVFVLAFAFKIPFEVIFPFIIGLISGLSYGLSYMDLTPLELPA
GFFLSIGTIGPSATGMIGFNGQIFVWAGILLGCLLLAAGGRKEAH"
trNA complement(1508..1579)
/product="tRNA-Glu"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 69.21"
complement(1585..1675)
/product="tRNA-Ser"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 63.76"
complement(1712..1784)
/product="tRNA-Asn"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 85.46"
complement(1791..1864)
/product="tRNA-Ile"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 92.91"
complement(1891..1961)
/product="tRNA-Gly"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 73.1"
complement(1984..2056)
/product="tRNA-His"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 72.11"
complement(2083..2155)
/product="tRNA-Phe"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 79.39"
complement(2160..2235)
/product="tRNA-Asp"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 73.32"
complement(2240..2313)
/product="tRNA-Met"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 78.6"
complement(2356..2445)
/product="tRNA-Ser"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 66.89"
complement(2468..2541)
/product="tRNA-Met"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 90.88"
complement(2553..2626)
/product="tRNA-Met"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 77.07"
complement(2681..2753)
/product="tRNA-Ala"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 84.64"
complement(2773..2846)
/product="tRNA-Pro"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 86.85"
complement(2857..2930)
/product="tRNA-Arg"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 74.21"
complement(2940..3025)
/product="tRNA-Leu"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 66.88"
complement(3040..3114)
/product="tRNA-Gly"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 87.88"
complement(3130..3211)
/product="tRNA-Leu"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 69.82"
complement(3217..3289)
/product="tRNA-Lys"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 88.37"
complement(3330..3405)
/product="tRNA-Thr"
/notes="tRNA-Asn-SE vs 1.3 result - Cove score = 95.38"
complement(3414..3486)

```



```

Db 138781 TTAGAAATACCGCGCTGTATTACAAAAATTAGACGGTGATACACGTCGTGGGCGACGA 138722
QY 261 LeuSerileArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 138721 CTTTCCATCGTTCACTACAGAAACCACATCAATTCGTGCTACCGGTGAATAATG 138662
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 138661 GAAGCAATCGAAACGTTCCATCCCGATCGTATGCTTCAAGAATTCTCGGCATGGTGAT 138602
QY 301 ValLeuSerLeuIleGluLysAlaGlnbNAspValAspGlnGluLysAlaLysAspLeu 320
Db 138601 GTACTTCTCTTATTGAAAAAGACACAAACTGATGTAGATGACAAAAATGAAGCTATG 138542
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 138541 GAACAAAAATGAAAGACACACAGATGACCTTAGATGACTTCTTGACCAATTCACAAA 138482
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 138481 GTAAAAACAATGGGACCACTAGATGACTACTAAAAATGATGCCAGGGGCAACAAATG 138422
QY 361 LysGlyLeuAspLysLysMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 138421 AAAGCCCTCAGACACATGCAAGTAGATGATAACAACTCCGTCACATCGAAGCGGTAATT 138362
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 138361 AAATCCATGACCAAAACGAAAAAGATAATCCGGACATCATCAATCGGACGACGAGAAA 138302
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 138301 CGAATTGCTGTTGGAAGCGGACGCCAATTCAAGAAATCAATCGCCTCTAAACAATTT 138242
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyClyLysGlyLysGlyLys 440
Db 138241 GCTGAATGAAAAAATGATGAAGCAAAATGACTGGTGGAGGAAAGGTAAGAAAGGTAAA 138182
QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMetAsnLeuProPhe 455
Db 138181 -----AATCCATTGGCAATTTCAAAATGCCATTT 138152

RESULT 8
AX413016
LOCUS AX413016 319630 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 7 from Patent WO0228891.
ACCESSION AX413016
VERSION AX413016.1 GI:21445474
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
REFERENCE 1
AUTHORS Glaser, P. and Kunst, F.
TITLES Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 7 11-APR-2002;
PASTEUR Institut (FR)
FEATURES
Location/Qualifiers
1. 319630
/organism="Listeria innocua"
/db_xref="taxon:1642"
BASE COUNT 105207 a 55428 c 66726 g 92263 t 6 others
ORIGIN

```

## Alignment Scores:

```

Pred. No.: 7,12e-93 Length: 319630
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 6 Gaps: 1

```

US-09-943-108A-2 (1-455) x AX413016 (1-319630)

```

QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 187924 ATGSCATTGAGGACTACTCGAGAGACTCCAGAAACATAGACAAATTCGCGTAAA 187983
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 187984 GCGAAAGTAAACGAAGCTGACGTAAAGAAATGATGGTGAAGTTCGTCTTACTA 188043
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 188044 GAAGCCGATGTTAACTTTAAAGTCGTTAAACAATTTATTAACACATAGTAAGCGACGTGCA 188103
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlnValIleLysIleValGln 80
Db 188104 GTTGGCGGAGAGTATGAAAGAGCTACACCCGCTCAACAGTATTAATTAATGTTCAA 188163
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 188164 GAAGAATTAACAAGCCTCATGGCGGCGAAGAAAGTAAATCGGAACAGCTGATCGCCCG 188223
QY 101 ProThrValValMetValGlyLeuGlnGlyLysIleValGlyLysThrThrAlaGlyLys 120
Db 188224 CCAACCGTTATTATGATGTTTCAAGGAGCTGGTAAACACACCACTTCAGGAAA 188283
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 188284 CTCGCTAATTTATGCTAAAAAATATATCGTAAACCTTTACTAGTCGACGAGATATT 188343
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 188344 TACCGACTGCGCAATCAACAATTAAGAAACACTTGGCAACAATAGATAGTCCAGTA 188403
QY 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180
Db 188404 TTCTCTTAGCGCATCAAGTAAAGCCAGTAAATCGCAAAACAGCTATCGCTAAAGCA 188463
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 188464 AAAGAAGAACATTTAGATTATGTCATTATCGATACAGCTGGTGTCTTCATATCGAGAA 188523
QY 201 AlaLeuMetAsnGlnLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 188524 ACTCTGATGGCAATTAACAAAGTGAAGAAATCGCTACGCCAACTGAAATTTTACTT 188583
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 188584 GTAGTTGATTCATGACTGGGCAAGACCGCAGTAAATGTGCCCAAGCTTCACGCAACA 188643
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 188644 TTAGAANATTACCGCGCTGTGTTATTAACAAATTAGACGGTGATACAGTGGTGGGCGACGA 188703
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 188704 CTTTCCATCCGTTTCAGTCAGCAAGAAACCAATCAAAATTCGTTCTACCGGTTGAAAAATG 188763
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 188764 GAAGCAATCGAAACGTTCCATCCCGATCGTATGCTTCAAGAAATTCGCGCATGGGTGAT 188823
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 188824 GTACTTCTCTTATTGAAAAAGACACAACTGATGTAGATGACAAAAATGAAGCTATG 188883
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 188884 GAACAAAAAATGAAGACACACATGACCTTAGATGACTTCTTGGACCAATTCGCAACA 188943
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 188944 GTAAACAAATGGGACCACTAGATGAATCTACTAAAAATGATCCAGGGGCAACAAATG 189003

```

```

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnLeuAspHisIleLysAlaIle 380
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189004 AAAGCCTCGACAACTGACAGTAGATGATRAACAACTCGGTACATCGAAGCGATATT 189063

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189064 AAATCCATGACCAAAACGAAAAAGATAATCCGGACATCATCAATCGCAGAGAAGAAA 189123

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189124 CGAATTCGTCGGAAGCGGAGCCCAATTCAAAGAAATCAATCGCTCTTAACAAATTT 189183

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189184 GCTGAATGAAAAAATGATGNAGCAATGACTGGTGGAGGAAAAGGTPAAGAGGTAAA 189243

QY 441 ArgAsnGlnMetGlnMetLeuLysGlyMetAsnLeuProPhe 455
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189244 -----AATCCATTGGCAATTTCAAAATGCCATTT 189273

RESULT 9
LOCUS AX417046 349980 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 4037 from Patent WO0228891.
ACCESSION AX417046
VERSION AX417046.1 GI:21449656
KEYWORDS
SOURCE
ORGANISM
Listeria innocua.
Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS
Glaser, P. and Kunst, F.
TITLE
Listeria innocua genome and applications
JOURNAL
Patent: WO 0228891-A 4037 11-APR-2002;
Pasteur Institut (FR)
FEATURES
Location/Qualifiers
1..349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/notes="seq 2058, original length: 3,011,208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"
BASE COUNT 101055 a 72969 c 60688 g 115268 t
ORIGIN

Alignment Scores:
Pred. No.: 7,89e-93 Length: 349980
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 6 Gaps: 1

US-09-943-108a-2 (1-455) x AX417046 (1-349980)

QY 1 MetaLapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgLys 20
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129501 AFGGCATTGAAGACTAGCTGAAGACTCCCAAGAAACAAATGAACAAATTCGCGGTAAA 129442

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129441 GGGAAAGTAAACGAAGCTGACGTAAAGAAATGATGCGTGAAGTTCGTCTTTACTA 129382

QY 41 GlnAlaAspValAsnPheLysValLysGluPheIleLysThrValSerLysArgAla 60
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129381 GAAGCCGATGTTAACTTAAAGTCGTTAAACAATTTATTAAACAGTAAGCAAGCTGCA 129322

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129321 GTTGGCGCGAGCTAATGAAGSCTTAACACCGGTACACAGTAATTAATAATCGTTCAA 129262

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129261 GAAGAATTAACAAGCTCTCAGGGGGGGAAGAAATAAATCGGAACAGCTGATCGCGCC 129202

QY 101 ProThrValValMetMetValGlyLeuGlnGlyValAsnLysThrThrAlaGlyLys 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129201 CCAACCGCTTATATGATGTTAGGTAGGTTCAGAGGAGCTGGTAAACCAACCACTTCAGGAAA 129142

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129141 CTCGTAAATTTATACGTAAATAATAATATCGTAAACCTTTACTAGTCGACAGATATT 129082

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129081 TACCGACTCGACCAATCAAAACAATAGAACACTTGGCAAAACAATTAGATATGCCAGTA 129022

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129021 TTTTCTCTAGCGCATCAGTAGCCAGTAGAATCGCAACAAAGCTATCGCTAAAGCA 128962

QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128961 AAAGAAGAATTTAGATTATGTCATTATGATACAGCTGGTCTTCATATCGACGAA 128902

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGlnIleAlaLysProAsnGlnIleMetLeu 220
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128901 ACTCTGATGCGCAATTAACAAAGTAAAGAAATCGCTACGCCCACTGAAATTTTACTT 128842

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128841 GTAGTTGATTCATGACTGGCGACAGACGAGTAATGTGGCCCAAGCTTCAACGACAAA 128782

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128781 TTAGAAATTAACCGCGCTTGTATTAAACAAATTAAGCGGTGATACACAGCTGGTGGCGACGA 128722

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128721 CTTTCCATCGTTCAGTCAAGAGAAAACCAATCAATTCGTTGCTACCGGTGAAAAAATG 128662

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128661 GAACCAATCAAGACGTTCCTATCCGATCGTATGCTTCAGAAATTCGCGCATGGGTGAT 128602

QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysLeu 320
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128601 GTACTTTCTCTTATTGAAAAAGCACAAACTGATGATGACAGAAAAAATGAAAGCTATG 128542

QY 321 GlnLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128541 GAACAAAAATGAAGACAGACAGCATGACCTTAGTAGACTTCTTGACCACTTCACAAACA 128482

QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128481 GTAAACAAATGGGACCACCTAGATGAACACTACTAAATAATGATGCCAGGGGCAACAAATG 128422

QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGlnLysGlnIleAspHisIleLysAlaIle 380
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128421 AAAGCGCTCGACCAACATGCAAGTAGATGATAAACAACCTGGTCACATCGAAGCGATATT 128362

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128361 AAATCCATGACCAAAACCAACCAAGATAATCCGGACATCATCAATCGCAGAGAGAAA 128302

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128301 CGAATTCGTCGGAAGCGGACGCCCAATTCAGAAATCAATCGCTCTTAACAAATTT 128242

QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```





Qy	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
Db	1491	TTTTCTATTTCGGCTGTCACAAACACCCCATTAGTTTGAGGTTTGGCGAAGAGCTT	1550
Qy	281	ASPGlyLeuLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
Db	1551	GATCGCTAGAGCCGTTCCATCTCTGAACGCGATGGCATCAAGGATTCTCGCANGGCGAC	1610
Qy	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu	320
Db	1611	GTCTGACATTGATTGAAAAAGCAGCGCGATGATGAAGACAGCAAGCCAAAGAGCTG	1670
Qy	321	GluLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln	340
Db	1671	GAACAAAAATCAGAACATGAGCTTCACATTGGACGATTTTCGGAGCAGCTCGGGCAA	1730
Qy	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
Db	1731	GTGAGAAACATGGGCGCCTTGATGAGCTCTGCAAAATGATCGCGGTGCAGGTAAATG	1790
Qy	361	LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIlelle	380
Db	1791	ARGGGCTGAACAAATCCAAAGTTGATGAAAAACAGCTGAATCATGTGGAAGCAATCATC	1850
Qy	381	GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys	400
Db	1851	AAATCAATGACTGTTCTTGAAAAGACAGCGCGATATATCAATGCGCCGCGGAAG	1910
Qy	401	ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe	420
Db	1911	CGGATTGCAAAAGGAGCGGACATCGTACAGGAAGTCAACCGTCTGCTTAAGCAGTTT	1970
Qy	421	AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys	440
Db	1971	GATGAATGAAAAAATGATGAAGACAGATGACAAACATGTCGCAAGCAAGAAA-----	2024
Qy	441	ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe	455
Db	2025	-----AAAGGTTTAGCTACCTTTT 2045	
RESULT 11			
BSUB0009			
LOCUS	BSUB0009	208780 bp	DNA linear BCT 26-NOV-1997
DEFINITION	Bacillus subtilis complete genome (section 9 of 21): from 1598421 to 1807200.		
ACCESSION	Z99112	AL009126	
VERSION	Z99112.1	GI:2633902	
KEYWORDS	Bacillus subtilis.		
SOURCE	Bacillus subtilis.		
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
REFERENCE	1 (bases 1 to 208780)		
AUTHORS	Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G., Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borcheret,S., Borriss,R., Bourllet,L., Brans,A., Braun,M., Brignell,S.C., Bron,S., Brouillet,S., Bruchi,C.V., Caldwell,B., Capuano,V., Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J., Daniel,R.A., Deniset,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D., Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E., Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A., Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Gollightly,E.J., Grandi,G., Guisepi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R., Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hulio,M.F., Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y., Kliner-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P., Koningsstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A., Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H., Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P., Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M., Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M., Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P., Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M., Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,		

TITLE		The complete genome sequence of the gram-positive bacterium	
MEDLINE		Bacillus subtilis	
JOURNAL		Nature 390 (6657), 249-256 (1997)	
PUBMED		98044033	
REFERENCE		2 (bases 1 to 208780)	
AUTHORS		Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.	
TITLE		Direct Submission	
JOURNAL		Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48	
FEATURES		Location/Qualifiers	
SOURCE		1. 208780	
		/organism="Bacillus subtilis"	
		/strain="168"	
		/db_xref="taxon:1423"	
		258..4588	
		/gene="bpr"	
		258..4559	
		/gene="bpr"	
		/EC_number="3.4.21.-"	
		/note="alternate gene name: bpf"	
		/codon_start=1	
		/transl_table=11	
		/product="bacillopeptidase F"	
		/protein_id="Csb13404.1"	
		/db_xref="GI:2633903"	
		/db_xref="SWISS-PROT:P16397"	
		/translation="MRKKTNKLISSTVTSWISSLLFFGAAGASSKVTSPVKKELO	
		SAESIONKISSLLKSKFKKTKTLFKFDLANPKAAKAAKAKSKLSAAKTEY	
		QKSAVYSSLKVTADESQDVLKYLNTQKQGNADQIHYYVVGIVAHASKFVWEKY	
		VQFPEVKVLNPKOLFSSPFNMKKAQAKATDGVENVNDQIDAPKAWALGYDG	
		TGTVAISDTGVWNNHPALKKEKYGNPENFENMWNVDYAVAGEVYDDLAHGT	
		HVTGMVGEEDPTNQIGVAPKAWIAKFAFSEGGTADILEAGEWLAPEAGNP	
		HPMAPDVNNSWGGSGLDWYDMVNAARAADIPFESAGNTDLFIPGPGSIANGP	
		ANYPESFATGATDINKLADSLQGPSYDEIKPEISAPGNIRSSVPQGYEDGNG	
		TSMAGPHVSAVALLKQANASILDVEMEDILTSTAEPLDSTFDPSPNGYGHGLVNA	
		FDVASVTDGLKAGQVSVGGDDQEPYVQHEKYTEAYEGGSLPLTLTADNVSVTS	
		VKLSKLDQGEWTEITAKRISGDHLKGTQAEIPDIKPKLSIKWMIHDFGVVSSD	
		YVDVTKPSTAGYKQDFETAPGGVAGSTNNNHWGVPSTGPTAASGEKVTGNT	
		GNTANANLNMPPIKAPDSGSLFQFKSNHLEDDFDYGVFVLPGEKNWEQAGV	
		YNGTSSWDEEDLSAYKGNIOVMFNLSDESIAKEGWIDDVVLSDKSGAGTKVKY	
		NKLGYEPKQKQKPKVPKAKPSANTAVKHONKAIOPQLPIKAQVSVYETGKSTY	
		SDASTQGYLIKHKAGDYILMEANYGYQSKTKYSLKTDQTQANTLBEAKKGLKGT	
		VINTKTEPTGASVTVVDEAAVEAPMTNDKGEIMLEAYEGAITKVAAPSYSEFS	
		VEIKRDVETAKLPFVGPGEIAYDDGTANANSYFAAGNWAYMTLADGDKGML	
		TGLFLRWDETPDFCGTEFFVEYDATGDKGAPKKIAGPENAALRNGWTKVDLS	
		SKGLMDQGFYLYVYQSKPDYSPGLAMDETQNSGRNWOYIDGKQWPGKADGNTMI	
		RALVDYEAAPVEITPTSTDKTSYTKNDSVTVKGNASPGTTVHIYNGEKEAGTKAADGT	
		FHAGILINGENELATATSTONGTDDASSPTVTLQDKEPLTLNPKDGGKINKETIL	
		TVYKAVSDNLKDKVKNKKAIVADGYSARILLRNGREIKVITATDLAGNKTATKTV	
		IDVNFKPVISGLIPGEDKNLKAESVKIAFSAEDLDATFTTIRMPLTNARASVQNA	
		ELPLREISFGRYEGYWTATSSIKAKGAKVEIVRDDYDGNTEKTKANGKNNTEN"	
		4571..4588	
		/gene="bpr"	
		4754..5683	
		/gene="spoIIA"	
		4754..5683	
		/gene="spoIIA"	
		/function="processing of pro-sigma-E (SpoIIG) to active	
		sigma-E (stage II sporulation)"	
		/codon_start=1	
terminator			
gene			
CDS			

```

/transl_table=11
/product="protease"
/protein_id="CAB13405.1"
/db_xref="GI:2633904"
/db_xref="SWISS-PROT:P13801"
/translation="MKYILDVTLNFCFDALLLLTAFILKRHKVKKRLVGGAFIOS
STVLMTTPSPIVEHPAGKLAFSVIVVTFGPKRPFPPFLSFYFAFLMGGOI
IGASLQSNISQVGMVMTNQTGCDPISWLFYGGFPALNPFSSKRRIEDLEIKNIO
YERVSADQIGSTQHLVRLGIDSGNQLYDELTPTPVNIIYIDKLEPIFGAETNIIR
NDPUEALEQDUSFRFLDKMLRILPIRGVGQGNFLCVRKPDHVTIMTKEMIADXC
LIGSTYKLSADGEFDAILHPKMLSKRAKVS"
gene
5746..6465
CDS
/gene="sigE"
5746..6465
/function="early mother cell-specific gene expression"
/note="alternate gene name: spoIIGB"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma-29
factor (sigma-E)"
/protein_id="CAB13406.1"
/db_xref="GI:2633905"
/db_xref="SWISS-PROT:P06222"
/translation="MKKILRLTHLWYLLMKLGKSDVEVYIGGSEALPPPLSKDEE
OVLMLKPGQOARAILERILRLVVIYARKFENTINIEDLISIGTIGLIKAVNPF
NPEKILATVASCATENETMLVLRNNKIRSEVSEYEPINLDWGNELLISDVLGPD
DOLTKDTEANVDKLLKALQNEREKQIMELFGLVGEENEKQKDVADMGISQS
YISRLERKRIIRKREKFNKV"
gene
6605..7387
CDS
/gene="sigG"
6605..7387
/function="sigG"
/note="late forespore-specific gene expression"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma factor
(sigma-G)"
/protein_id="CAB13407.1"
/db_xref="GI:2633906"
/db_xref="SWISS-PROT:P19940"
/translation="MNSRNVKEICGVDTSKLPVLKNEMRKLFQLQDEGDSAREKLV
NGNRLVSIQRNENGEYVDLFGVCGIGLMKSIDNFDLSHNKFSYVAPPIGE
IRRYLDNNRPSRSLDIAYKQVRELTISNTPSKPTAEDIAKLVLEPHEITVA
LDAIODPVSLEPIYNDGGDPIYVMDQISDERNTDSQWIBELAKEGMRLENDREKMI
LKRFFQCKTOMEVAEEIGISQAQVSRLEKAIKQMKNIHQ"
gene
7335..8496
CDS
/gene="yima"
7335..8329
/function="unknown"
/note="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB13408.1"
/db_xref="GI:2633907"
/db_xref="SPTREMBL:O31723"
/translation="MILQDNYSLKRNKGWILKDIHWKVEKENVLYLGNAGKTL
LNMCSYFPSPSGEMRYLGHFKTELGKLRKIGLVSAALQKLYPADSAFETALS
GAYASIGLYFPSPKREKALGLEDGAIYADRRYETLSQEKORALLARALMADP
ELLIDPVGTEKMLDITITIANKENAPSILYVTHAEELTFVDFKALLIKQ
GEVFGSGKMLDQILSAFDTPIHVLNQDRFFVTRAEPITNA"
terminator
8472..8496
/gene="yima"
8531..9811
gene
8531..9811
CDS
/gene="yimb"
8531..9811
/function="unknown"
/note="similar to acetylornithine deacetylase"
/codon_start=1
/transl_table=11

```

```

/protein_id="CAB13409.1"
/db_xref="GI:2633908"
/db_xref="SPTREMBL:O31724"
/translation="MDQIYSLQKVEEHEKEELIQLAKTL-SYOTPPAPPARTNEGIOS
WIQYLNELGFSIDKWDVPGDPNVVGRLAGTSADYSLIINGHVDVAVKDEEWK
HDPHPTEKNGLLIGRGSMDKMGACVLFVAKLIREASIELPGDLILQSVIGEEVK
AGTLECCCKRGYHADRAIVADTSDMHIOGGGVITGWIEFKSOTPHGTGRNMHAGG
GTGCASIRKMAKLIAGLELERHWSIMKSYFGFPGTNTNPVAVIGSGHRAFIAD
CRUMITVHFYFNETHDQAAELEDYVNLRSDDIWLRENRPVFRWGGSSMIEDRGEIF
PALEVDHPGVLTAUTHSKVKECPIDVSQSVTDGGLYDAGIFCVIYFGDLHN
AHSVNEKVSIEQOLVEYTRKILIDFTIISWCSRKEQ"
gene
9894..10139
CDS
/gene="yImc"
9894..10139
/function="yImc"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB13410.1"
/db_xref="GI:2633909"
/db_xref="SPTREMBL:O31725"
/translation="MISTSEFQVXDVNVNNGKLGISGIDIDINVTGKTQAILGNN"

Alignment Scores:
Pred. No.:      2,38e-89      Length:      208780
Score:          1647.00      Matches:      319
Percent Similarity: 84.62%      Conservative: 66
Best Local Similarity: 70.11%      Mismatches:   60
Query Match:      72.43%      Indels:       10
DB:              1          Gaps:         1

US-09-943-108a-2 (1-455) x BSUB00009 (1-208780)

QY   1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB  73153 ATGGCATTTTGAAGAGTATGACGACGCGCTGCACAGACAGATTTCCTAAATCCGCGGAA 73212
QY   21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
DB  73213 GGGAAAGTCAGCGACACAGATGTAAGACATGATGCGTGAGGTCGCTTGGCTGCTT 73272
QY   41 GluAlaAspValAsnPheLysValValLysGluPheLleLysThrValSerGluArgAla 60
DB  73273 GAGGCTGACGTTAACTTTAAAGTAGTCAAGGATTTGTCAAAAAAGTAAGTAACGCGT 73332
QY   61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
DB  73333 GTAGGCGAAGACGCTCATGAAAGAGTCTGACGCGCGCGACAGGTCATTAAGTGTTCAA 73392
QY   81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
DB  73393 GAGGAAGTCACTGAGCTGATGGGCGGCGAAGAGAGCAAAATCGCGCGTCGCAAAAGGCG 73452
QY   101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
DB  73453 CCGACTGTTATTATGATGTCGCTCCAAAGTGGCGGTAAACACACACACACGCGGTAAG 73512
QY   121 LeuAlaLeuLeuMetArgLysLysTyraAsnLysLysProMetLeuValAlaAlaAspIle 140
DB  73513 CTTCGGAATCTGCTCGGCAAAACATATCGCAACCGATGCTGCTGCTGCGGATATT 73572
QY   141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
DB  73573 TACGCGCCAGCGCAATTAAGCAGCTGGAAACACATCGGCAACAGCTTGATATGCTGTT 73632
QY   161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
DB  73633 TTCTCTCTTTGGCGATCAGTCTCTGTAGAAATAGCTAAACAGGCTATTGAGAAGCC 73692
QY   181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
DB  73693 AAGGAAGAACATTATGATACGTCACTTTTGATACGGAGGCGCTTGCAATATCGACCAT 73752

```



FEATURES  
source

```
Location/Qualifiers
1. .303249
/organism="Bacillus halodurans"
/db_xref="taxon.86665"
/note="alkaliphile"
104. .1102
/gene="BH2267"
104. .1102
/gene="BH2267"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein"
/protein_id="BAH05986.1"
/db_xref="GI:10174887"
/transl_table="MKNHLHTIMEDWKLGSVALMKKGEDIPFIASLGRANRAERIPNE
HHTFRGASGKLFATAICQLVEAGKLSFDTPLSDWLDAFPNVTIHLHTHTSGYP
DYFDEITDDFEDLWKDVMYHLRLKDFLFLCHAPMKFPFGHFRNINAGFTILLGL
VYESVGVTQEVYEANVQFRAGMHESGIFAFDLPAKTALGIDLEPGSKWTLYSL
PVIIGDSGAYVAEDMMKMLALMRHELLNETYTKLLTPPHVCEDDDYGYGWNK
QODGAIKXYHVMGYDFGVCFHSAFYPTSGIVVVCANQSSGAYDVMAAIEALFSEA"
/gene="BH2268"
/complement(1146. .1469)
/gene="BH2268"
/complement(1146. .1469)
/gene="BH2268"
/note="BH2268
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAH05987.1"
/db_xref="GI:10174888"
/transl_table="MKVITLCGSTGTFKRFQFQASAYLTLLKGHVLSVGFFQSGIET
TEBQVQMFENIHRKTIEMSGDEIFVLVDVGYIGRSTRKEIFATSKGMPIRYSESLA
STTFS"
/complement(1636. .2373)
/gene="BH2269"
/complement(1636. .2373)
/gene="BH2269"
/note="BH2269
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAH05988.1"
/db_xref="GI:10174889"
/transl_table="MYKKVLLIVGVFALSIGLMACNQSQEVEEKNIDGEASGIPSI
BEIDINLPDPVATVNGESIDKETVYVLFQQMTMAQYG:SLGEEGKQMSLIEP
OTIDOLNEQLQAAATEKEIEVSEEDRELATFVAOFSSEALMEALSGGSMDE
LREIEHVQSQKYVEETEAINVTESEIQARVEEKQYITDEFLTPEELQSNIEOR
LIAKKEQEIENVLEKLRBEGDITVHI"
/complement(2582. .3367)
/gene="BH2270"
/complement(2582. .3367)
/gene="BH2270"
/codon_start=1
/transl_table=11
/product="ABC transporter (ATP-binding protein)"
/protein_id="BAH05989.1"
/db_xref="GI:10174890"
/transl_table="MADILQVEQVTLRAKKKPILOQISQVQEQGHEWAIVGLNSGKT
SLKLIITGYEMPTGEHVAIVNKRGHVP:IOEVKRKIGWYSMSLDDRPFTTKGQDTVLEI
VLSGLHGTGVYPLGEADVAKARAATETGLTLHLDERFLTLSOGEKRAFTARAVY
AKPEFLIDBETFTGLDLIAREQLLOTLESVATPKGAPTELYVTRHPBEIQPIETHVM
LADGHVTAQGRKDEVLTKEMITAFAGCSLDVIAQSGRIYWKPM"
/complement(3378. .4310)
/gene="BH2271"
/complement(3378. .4310)
/gene="BH2271"
/codon_start=1
/transl_table=11
/product="dipeptidase"
/protein_id="BAH05990.1"
/db_xref="GI:10174891"
/transl_table="MEIFDAHCDALLKLEDPAREYTDOPDAIEANAYVLMRGVRAQC
FAIVFGNIFFDKQFVALQQLFHEKILKKHPQSFIERWDDLLTLDDAQIGAILT
```

## gene

```
LEGADPFGDDPTKLTLYRLGYRSLGLTWNQANLCADGVGPEPRGGGLTEWGAUVVSLN
NEHVLTVDSHLSERGFMDVIEYAQFPVAPHSNKALCKHRNLTAOLKALIEAGGF
IGIYVHPHFYLTGTTDASIDDVKKHIEHCISLGQKQKHEVGRGSDFDGIFCVGLEHAGK
FCWFMDELLARIDEELVGRFAWDNFKATP:LSYV"
4406. .4840
/gene="BH2272"
4406. .4840
/gene="BH2272"
/codon_start=1
/transl_table=11
/product="heat shock protein class I (low molecular
weight)"
/protein_id="BAB05991.1"
/db_xref="GI:10174892"
/transl_table="MSRESGSKLPVNYQDFLRSDDDFSKSFQRLHDNMGFAPSTP
IRQKEQGEQIIEAELPGVSKNQICLDVFFQYVIRIHEEVELEMEKDEKGIYSKSL
SQAKERVIPPPYVQSGDIKARLNRGLLTAIPNHRKRINIE"
4864. .5193
/gene="BH2273"
4864. .5193
/gene="BH2273"
/note="BH2273
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB05992.1"
/db_xref="GI:10174893"
/transl_table="MYNLHPQRTWSQAGYVQQPFQHSTWPNHPQRTSSGQWMTSTG
FSPPFPYNYFSQQKPRASPKAAFTNQDGTDFINKTFQTVQVVKTVNQVTPLYKS
VGSIFIK"
/complement(5202. .5510)
/gene="BH2274"
/complement(5202. .5510)
/gene="BH2274"
/note="BH2274
unknown"
US-09-943-108a-2 (1-455) x AP001515 (1-303249)
Qy 1 MetaAlpheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 210434 ATGCGCTTTGAAAGTTTAGCGGAGCGGCTGCAAGATACGTTGAAGAAAATCGCGGAAA 210375
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 210374 GGAAGAGTAAGTGAACAGACGCTPAAAGAAATGATCGGTGAAGTTCTGCTTCCTCTTA 210315
Qy 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 210314 GAAGCGGATGTTAAACTTTAAAGTGTGAACCAATCATCGCAATGTGAAGGAAAAGCC 210255
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 210254 CTCGGTCAAGAGTTCATGAAAGGCTGACACCTGGACACAGGTCATCAAGTCGTCAAT 210195
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 210194 GAGAGCTGACGGCGCTGATGGGAGCGGAGCAAGAAAATGCGGTGCGGACAGACCG 210135
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 210134 CCAACGGTGTGATGTCGTCCTTCAAGGGGCTGGGAAAACGACGACGCGGCGAAA 210075
Qy 121 LeuAlaLeuMetMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
||||| ::::::::::::::::::::: ||::::::::::|
```

## Alignment Scores:

Pred. No.:	2,43e-85	Length:	303249
Score:	1584.00	Matches:	307
Percent Similarity:	83.33%	Conservative:	73
Best Local Similarity:	67.32%	Mismatches:	70
Query Match:	69.66%	Indels:	6
DB:	1	Gaps:	2

Db 210074 CTTGCCAACCACTTTGCGGAAAAAGCACAAATCGGAACCGGATTCTCGTCGCTGCCGATATT 210015  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
Db 210014 TATCTCGCGTGCAGTAAACAGCTTGAGACATTAGTAAACAATTAACATCGCTCTG 209955  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
Db 209954 TTTACCTTGGTGATCAAGTAAGCCCTGTGAAATGCGGAAACAGCGATCGAGCGGCG 209895  
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
Db 209894 AAAGAAGACATCACAGCTATGTGCTCATCATGATGAGTGGAGCTTTCACATCATGAA 209835  
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
Db 209834 GAGCTGATGGATGAACGTGCAACAGTAAGTAAGAGTGGCCAGCCAGCATGAATCTTGCTC 209775  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
Db 209774 GTCTGTGATGCGATCACAGGCGAAGATGCGTTAATGTGCGGGAAGCTTCCATGAGCAG 209715  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260  
Db 209714 CTCGATGTTTCGCGCTCGTGTTCACGAAGCTGGATGGGATATCATGAGCGGAGCTGCC 209655  
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
Db 209654 CTCTCGGTTAAGCGCGTTACGAACGCGCATCAATTTGCCGGGATGGGTGAAGAAT 209595  
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
Db 209594 GACCAGCTCGAACCGTCCACCCAGAGCGAATGCGTTCGCGGATCTCTGGGATGGGGAT 209535  
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
Db 209534 GTGCTGACACTCATTTGAAAAGCACCAATCCACTGTCGATGAAGAAAAGCGCGTGAACCT 209475  
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
Db 209474 GAGAAAAAATGCGTACAAATGGACTTACGTTTCGATGATTTCTTGAGGCACTCCACCA 209415  
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
Db 209414 GTGCGGAGCATGGGACCACTCGATGAATGCTCGGAATGATGCCAGGATGAACAAACA 209355  
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
Db 209354 AAAGGCTTGAAAAAATTTGCAAGTGGACGAGAACCAATTAAGCCGCTTGAGCGGATGTC 209295  
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
Db 209294 CGTTCCATGACGAAGCAGAGAGAACACAGATCTTAGCATCAATATGCGAGTGTGTCGA 209235  
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
Db 209234 CGGATCGCGAAAGGAAGCGGCGCATTCGAAGATGTCACCGCTGTTAAAGCAGTTT 209175  
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGly---GlyGlyLysGlyLysGly 439  
Db 209174 GAAGATGAAAGAGATGATGAAGCAATGACAAACATGCGCAACATGCGCAAGCAAGCA 209115  
QY 440 LysArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
Db 209114 GGGAAAGGC-----TTTGGAGGCTTTAAATGCGGTTT 209082

RESULT 13  
AX433432  
LOCUS AX433432 1136 bp DNA linear Pat 28-JUN-2002  
DEFINITION Sequence 1847 from Patent WO0229113.  
ACCESSION AX433432  
VERSION AX433432.1 GI:21658236  
KEYWORDS

SOURCE Bacillus licheniformis.  
ORGANISM Bacillus licheniformis  
REFERENCE 1  
AUTHORS Berka,R. and Clausen,I.G.  
TITLE Methods for monitoring multiple gene expression  
JOURNAL Patent: WO 0229113-A 1847 11-APR-2002;  
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)  
FEATURES  
source location/Qualifiers  
1..1136  
/organism="Bacillus licheniformis"  
/db\_xref="taxon:1402"  
BASE COUNT 338 a 263 c 326 g 209 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,68e-77 Length: 1136  
Score: 1402.00 Matches: 270  
Percent Similarity: 87.30% Conservative: 60  
Best Local Similarity: 71.43% Mismatches: 48  
Query Match: 61.65% Indels: 0  
DB: 6 Gaps: 0  
US-09-943-108A-2 (1-455) x AX433432 (1-1136)  
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
Db 1 ARGGATTCGAGAGATTAGCGACCGACTGACGAAACAGATCTCAAAAATCCCGGAAAA 60  
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
Db 61 GGAAGGTAAAGGACGACGATCAAGAGATGATGCGGAAGTCGCTTTAGCGCTCTT 120  
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60  
Db 121 GAAGCGAGCGTCAATTTAAAGTCGTGAAAGATTTTGTGAAAAAGGTAAAGCAACGAGCT 180  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
Db 181 GTCCGACAGAGCTTTGAAACACCTGACGCCCGGACAGCAGGTCATTAAAGTGGTAAAA 240  
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
Db 241 GAAGAGCTGACGAGCTGATGGCGGCGAGGAAAGCAAGATTGCGCTGCAAGAGCTCG 300  
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
Db 301 CGAACCGTCATCATGATGGTCTCTGCAAGGGGCGGTAAACGAGCTCGACCGGAAAG 360  
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
Db 361 CTTGCCACCTGCTTCGCAAAAACATAACCGCAATCCGCTGCTCGTGGCAGCAGACATC 420  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
Db 421 TACCGCGCGGCGGATCAACACAGCTGGAAGCGCTCGCAAGCAGCTCGATATGCGGTA 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
Db 481 TTTTCACTGGGTGACAGGTCAGTCGCGGTGGAGATTGCCAAGCAGCGCATPCGAAAAGCG 540  
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
Db 541 AAAGAAGATCATCAGCTACGTTCTCATCATGATGCGGCGGACGCTTCATATGATGNA 600  
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
Db 601 GAGCTGATGATGAGCTTGACAGGTAAAAAGAACCGCTCAGCGGAGGAGAGATTTTCCTT 660  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
Db 661 GTCTGTGACTCCATGACGGGTCAAGCGCGGTCAATGTTGCAAAAGCTTTTACGAACAG 720





```

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 8487 CTTTCATATAGAGAAATGACTGGAACCTATAAAATATGTGGTTTGGTGAGAAGATG 8546
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 8547 AATGATATAGATCTTTTCATCTCATAGATGCGATCAGCAATACTTGGTATGGTGAT 8606
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 8607 GTTTTACACTTATAGAAAGCTCAGTCTCTATTGATGAGAGCAAGCAAAAGACCTT 8666
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 8667 GGTGACAGATGCTCAGCAGCAATTTAAATTTGATGATTTCTTCACGCTTTTGACAA 8726
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLys--- 359
Db 8727 ATGAAAAGTTAGGACCTATAGTAACTTCTTGAATGCTTCCAGGTTTAAATAGTAGT 8786
QY 360 ---MetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAla 378
Db 8787 ATGCTTATAGGTGTGTAT---TTGTCAAAAACGAGAGGTGAATGAAAAGTATGAGGCC 8843
QY 379 IleIleGlnSerMetThrProAlaGluArgAsnAsnProAsp-----ThrLeuAsnVal 396
Db 8844 ATTATAAATCGATGACTCTAGGAGAGAGAAAATCCATCTTATTAATCTTACAGCT 8903
QY 397 SerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGlnValAsnArgLeu 416
Db 8904 TCGAGGAAAAGAGATATGCTCTTGATCAGGTACTAGCTTCAAGAAAGTAAACAAAATA 8963
QY 417 MetLysGlnPheAsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGly 436
Db 8964 CTTAAAACTTCGAGCAATGAGAGATGATGAGCGATTAAAGGAAAATGAAGTTTCT 9023
QY 437 LysLysGlyArgAsnGlnMetGlnAsnMetLysLysGlyMetAsnLeuProPhe 455
Db 9024 AAAAAAGG-----TTATTGGTGGGAAAATGCCTTT 9056

RESULT 15
AP003191/c
LOCUS AP003191 296750 bp DNA linear BCT 10-JUL-2002
DEFINITION Clostridium perfringens str. 13 DNA, complete genome, section 7/10.
ACCESSION AP003191 BA000016
VERSION AP003191.2 GI:18146729
KEYWORDS
SOURCE Clostridium perfringens str. 13 (strain:13) DNA.
ORGANISM Clostridium perfringens str. 13
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
REFERENCE 1
AUTHORS Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A., Shiba,T., Ogasawara,N., Hattori,M., Kihara,S. and Hayashi,H.
TITLE Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
MEDLINE 21664373
PUBMED 11792842
REFERENCE 2 (bases 1 to 296750)
AUTHORS Shimizu,T.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical Sciences, University of Tsukuba, Department of Microbiology; 1-1-1 Tennohdai, Tsukuba, Ibaraki 305-8575, Japan
(E-mail:tschimizuend.tsukuba.ac.jp, Tel:81-298-53-3354, Fax:81-298-53-3354)
COMMENT On Jan 14, 2002 this sequence version replaced gi:18145205.
FEATURES
+ source
1. .296750
Location/Qualifiers
/organism="Clostridium perfringens str. 13"
/strain="13"

```

```

/db_xref="taxon:195102"
/note="anaerobic pathogen for gas gangrene"
complement(87..773)
/gene="biob"
complement(87..773)
/gene="biob"
/note="228 aa, similar to sp:BIOD_METJA DETHIOBIOTIN SYNTHETASE (EC 6.3.3.3) (DETHIOBIOTIN SYNTHASE) (DTB SYNTHETASE) (DTBS) from Methanococcus jannaschii (248 aa); 33% identity in 221 aa overlap CPE1543"
/codon_start=1
/transl_table=1
/product="dethiobiotin synthetase"
/protein_id="BAB81249.1"
/db_xref="GI:18145206"
/translation="MENGVIITGNDTKFTISGLILKLRREGRNAGYKAVLSGA IKERKGLIPDCEENKISGRYENVMVSYLSPHASEVEEVSIMKTKD YKSRDKYDFILCEGGIVCPISFEKKLMLEDIKEFNLOILVNSGLGTINVT LVSYLRNLGLKVKGIILNKFNKSDIIHRDNKKIKELTGNNISTVPKIEDIEKYL NELNEVLYGI"
complement(776..1735)
/gene="biob"
complement(776..1735)
/gene="biob"
/note="319 aa, similar to pir:E70384 biotin synthetase (EC 2.8.1.6) from Aquifex aeolicus (332 aa); 43.4% identity in 311 aa overlap CPE1544"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="BAB81250.1"
/db_xref="GI:18145207"
/translation="MDFILKMDKSLKNIKLTREGLRLNSMLELKEANNIRKEI HGDGIDGSIINGSGRGCEDCAQSKYHKTNISEYPLDYEKKYKAKENEDGV HRFSTVSGRLGEEFFERYTYNSLAKELINLCASHGINKESLIKRAKGYR HNLTSSRYDYKCTKTHSYERVKTIKNAEAGLEVCSGGIIGLGETILDRIDAIT LRELKSIPINVLSAIKGTQLQNMIPLENEILRTAVFRFINPEAKIRLAGGYLL KNFGNAFKAGANATITGNLLTTCGNKIKDDKRLIENIGMRIF"
complement(1754..2308)
/gene="biob"
complement(1754..2308)
/gene="biob"
/note="184 aa, similar to sp:BIQY_BACSH BIOY PROTEIN from Bacillus sphaericus (215 aa); 43.4% identity in 173 aa overlap. Putative N-terminal signal sequence and 4 putative transmembrane regions were found by PSORT."
/codon_start=1
/transl_table=1
/product="biotin synthase"
/protein_id="BAB81251.1"
/db_xref="GI:18145208"
/translation="MKMKYFDLPIPIFAALTAIGAIKIPIDIVFTQIYFFCALGA ILLGKALAAQIIYVLVGLGIPVQTGGGSYVQPTFGYILIGFIYAIIGKITE NKLTNIKFLAGLGLVILYLVFGVSYMYIYFNFLGDAMSLISAVSGIACLPSSD LLITFIISLVGSKVCPRLKRFDI"
complement(2694..3785)
/gene="CPE1545"
complement(2694..3785)
/gene="CPE1546"
/note="363 aa, similar to sp:YPUA_BACSU HYPOTHETICAL 31.3 KDA PROTEIN IN LYSA-PPTB INTERGENIC REGION (ORF19) from Bacillus subtilis (290 aa); 31% identity in 271 aa overlap. Putative N-terminal signal sequence was found by PSORT"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAB81252.1"
/db_xref="GI:18145209"
/translation="MKVLOGAIVSTFMFGMATPVFAOSEKTEAVVTLGANLPKSRLO

```



```

MLDAFGVKAHEVKIIDVTNQDIREQLGDTSKPIPASSQSISISSYVVVVKDKGINVTT
NNLEFVSGMIANALTSQVNADVADAPKVTGTAAAGLTKGFDAAGEELSLPK
KEAREISITNNLSNAKTQGLKDEAAVVDIKTDVLDKPKNDEBIGLYNN
VTNNILUFGQOQEGTIFKFSKINDIDYNGAKESLNMNDKIQQLKDKTGKQLEE
SGLLEKALNGIKNVLVDIKDFLVNFFSASEKVKDGTIDENGIVIKTGNNSDESKN
BESIQDPQTSQNDNNQENBQGNQTK"
complement(4141..7578)
/gene="spovk"
CDS
complement(4141..7578)
/gene="spovk"
/gene="1145 aa, similar to gp:AB015670.7 Bacillus sp.
genes for Cbase, Cbase, MBP and 15 ORFs, partial and
complete cds from Bacillus sp (549 aa); 45% identity in
522 aa overlap. Also similar to prf:2516401X stage V
sporulation protein k from Bacillus cereus
CPE1547"
/codon_start=1
/transl_table=11
/product="stage V sporulation protein k"
/protein_id="BAB81253.1"
/db_xref="GI:18145210"
/translation="MLQKNIFVERKDMVKNKEIENGFFKIEKKLNEEPIGOKTFNKL
CDYFKIESDSGILLIIVEEDIEFNKSVLKYFPEELNYKFKVNSKIDEIDLAAYF
NLGNAFLTDLYELKNDSCQLFKNTDKASEDILNLVSGISPNCSIMLADNYVKNK
FLVAREIDIDVYKNIICNNKFFIPLYNKKGEBLEIIDNEFLINKDKTLRKALSEK
DNKVYKKLLKINKVKELFDIDIIIGVDENPDVEKSGVCTYEDNFRKSGLSLSD
QYISKYCKPILINIIRKEAIEANSRLIYVEDDIEFCKSDNDEVLYLSKSPLEEVYR
YKLSIIGMKELMEIFDKLENNFKVKIREKLGKTKQISLNMIFAGNAGTKNAAR
ITFDYLVAGMIEKIFKEVSKADFIGEISNTARTNEIIFSAALGVLFIDEAYSIC
TDENKAGNIEVDALLKEDNRNDITVILAGYEKDMETFLFNOQLSKRPPIKFE
DNPSEMYETAVNIAKSGYRIAKAKAGLJELFKNQMGVSKDGNARFVRNIVENA
IWDASKLYDKERRAIDLEKDNFNKVSAKFLEEKLEIIGLEPKVEFRFSYKLI
VAQKRKSGVGNPKIBQNLNMFAGNPGCTGYTSIARLYAEMLSGKLLKVGQLYETDR
SVFSDIPGTSKRTKEKFEALGGVLFIDEAYTLANDSIGREAIETLLKLEIDISRE
VIVILAGTEEMENFTDVIAGLSRFLPMTKFDYNPNELEMAIKLBAKGLSKSN
IKAMNAIKYDNIHEDLQKLNLIENELKDFLGOYKLMIRKRLKGLFGVDLNKYM
NMVFTSGSGTKVLTNYSEMLYSGLIKAKNIVDQKYEPAMINSKMGVEDLANK
HYGLIYDKNFTTGEERYNEIVSDLIKFIDNNSRIVIVLGGIRGKMKDLITNEG
LNLRFYVWDFGNVNERELYDIASSTLVKKGFTLDEEAETALENAIGDIQKRWGELAL
KNGLMIKQFLDSLVVQSIRVCDEDFDINTNRIISEDILKSTEIPFKNNTTQNGSF"
complement(7634..9622)
/gene="fdh"
CDS
complement(7634..9622)
/gene="fdh"
/note="662 aa, similar to pir:S66017 formate dehydrogenase
homolog yVae from Bacillus subtilis (667 aa); 32.1%
identity in 667 aa overlap
CPE1548"
/codon_start=1
/transl_table=11
/product="probable formate dehydrogenase"
/protein_id="BAB81254.1"
/db_xref="GI:18145211"
/translation="MEVLSHGCTLDCCFCCKFNVYKEGSEILKIEGDKBHPPTKGLIC
KKGVAHLNRLNKHDRIVTPLLKNGVWEIEISFEDALIMKEKLEYTKVSKSSTLYY
SOYSGGVVLKIEDIEFNYGGVSRATGCPGWSGMRQAKYDFGDSVSNSEDMNSK
NIFLMGKNPANTIIHTMALNRAKNGSKLIIVDIPNTQSAKLGDHVAIKFETDQAL
AMAKAIISKGLDQDLINKYVLFQFYQKHLENFDLYLSDCEGIEIEDIEKLTLY
YCKRNSSTLYGMYKNGKNTIRADALGALTQIGVKGGVYANKVLSRILSDS
PFKSEVGENEFYVSNIEEPKYSLSVEDSNAPIKIMVIANSNLMNQLPNLR
LANSYDKYERKVCDFMDPTASKDCLPTCNTLSESDMVFSSMTNPYLIYNEKIIIE
PREKLDIYFFRELAKMNLKGYPSLSKYLSKVTEPLMSYKNDITLIDKNPFT
VYDDVAKENKFKTPSGKFELASKRALKEGSLTPYLSPIREKCFRLUHTNHSKSL
SSQHYIDYDEKAKYLLNENMRKFSLICGEYKVKLSKRTGTAICSLDNGVQDYVALM
YVWGMKGKGNPNFTESGISDMGGQITYNETPIETENI"
10033..11127
/gene="prs"
CDS
10033..11127
/gene="prs"
/note="364 aa, similar to pir:A71345 probable
phosphoribosyl pyrophosphate synthetase (prs) from
Treponema pallidum (421 aa); 32.3% identity in 375 aa

```

```

overlap
CPE1549"
/codon_start=1
/transl_table=11
/product="probable phosphoribosyl pyrophosphate
synthetase"
/protein_id="BAB81255.1"
/db_xref="GI:18145212"
/translation="MNEQNEGLGIITALESCTELGNADKLIQKRNCESEFLIKTDEI
RSGNGEKVPIQSVRGRIYIICDIGNYCTYXMGFENKNGKPDHGFQDIKRVSAI
RFGKARITVIMPLIYESRQRRKRGESLDICALQELERUGVDEVLTDFDHPNQNA
IPKSFENFYPTDYIVKSLIHNKDELDKELIVISPDGAMDRAIYSSVLGVQNG
LIFYKRRHSIRVANGKPIVKHYEMGRDVGDKVLIVDDMIASGESVLDIAKELGRNA
RNWYVATTFSEFTGELEKFNKFNFDGIKSVYSTNLTYIPEELNAPWFKAVDLSDFI
ARILNRLNYDKSVASYMDATNIIQRLDLEK"
complement(11177..12238)
/gene="CPE1550"
CDS
complement(11177..12238)
/gene="CPE1550"
/note="353 aa, no significant homology"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB81256.1"
Alignment Scores:
Pred. No.: 4,39e-73 Length: 296750
Score: 1382.00 Matches: 279
Percent Similarity: 78.88% Conservative: 72
Best Local Similarity: 62.70% Mismatches: 86
Query Match: 60.77% Indels: 8
DB: 1 Gaps: 5
US-09-943-108a-2 (1-455) x AP003191 (1-296750)
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuAlaThrMetGlnLysMetArgGlyLys 20
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 185046 ATGCGTTTTCATGATTAGCGTCTAAATTCACAGACACGCTTAAAGGTAAA 184987
QY 21 GlyLysLeuThrGluAlaAspLysLysLysLysLysLysLysLysLysLysLys 40
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184986 GGTAAAGTAAACAGAAAAGATATAAAGACCCATGAGAGAAGTAAAGCTTGCCTTTA 184927
QY 41 GluAlaAspValAsnPhelLysValLysGluPheLysThrValSerGluArgAla 60
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184926 GAGCAGATGTTAACTTCAGGTGTGTAAAGAGTTATATCAAAATGTTAAAGATATAATGT 184867
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysLys 80
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184866 GTTGTGAGGAGATATAATAGTTTAACTCCAGGTCAACAGGTTATTAAATAGTTAAT 184807
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysAsnMetSerAsnLysPro 100
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184806 GACGAATTAACAACCTCTTGGGAGAGAGCTGAGAGTAATAATAACTACTCAGACAAATGGA 184747
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184746 CCACAGTGTTCATGTTAGTGTAGTCTTCAAGTCCAGGTAAACAACTATGGCTGGGAAA 184687
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184686 CTTGCAATTCACCTTAAGAAAAGAAA---AACAAAAAGCCATTACTTGTGGCTTGTGATATA 184630
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184629 TATAGACTCCAGCTTAAGCAATATACAGGTTGTAGAAAAACAATAGATATACCACTA 184570
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184569 TTCCAATGGGAGATAAGGTAAAGCTGTAGATATGCCAAAGCTGCATAGACATGCA 184510
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisLysGlu 200
|||||...|||||...:|||||...:|||||...:|||||...:|||||...

```

Search completed: February 25, 2003, 02:59:00  
Job time : 4087 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 00:19:11 ; Search time 296 Seconds  
(without alignments)  
3461.687 Million cell updates/sec

Title: US-09-943-108a-2  
Perfect score: 2274  
Sequence: 1 MAFEGLSERLQATMQMRGK.....GKKGRNQMNLMKGNLFF 455

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xip  
-O=/cyn2\_1/USPTO.spool/US09943108/runat\_21022003\_151629\_1790/app\_query.fasta\_1.647  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPPCI=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09943108.ccn\_1.1.396 @runat\_21022003\_151629\_1790 -NCPU=6 -ICPU=3  
-NO\_XLIFY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2274	100.0	1368	20	AAx27221	S. aureus fff codi
2	2261	99.4	1368	23	AAx55331	Staphylococcus aur
3	2235	98.3	1365	23	AAx51783	Staphylococcus aur
4	2235	98.3	1368	23	AAx54393	Staphylococcus aur
5	1719.5	75.6	2256	24	ABQ70511	Listeria monocytog
6	1708.5	75.1	319630	24	ABQ67194	Listeria innocua c
7	1534	67.5	1014	24	ABN92495	Staphylococcus epi
8	1418.5	62.4	1416	23	AAx51308	Enterococcus faeca
9	1418.5	62.4	1434	23	AAx53127	Enterococcus faeca
10	1417.5	62.3	6729	20	AAx12874	Bacillus lichenifo
11	1402	61.7	1136	24	ABK74556	2CFE 28 coding seq
12	1312	57.7	1605	22	AAH90837	Streptococcus pneu
13	1307	57.5	1572	23	AAx55519	Streptococcus pneu
14	1306	57.4	1395	23	AAx55790	Streptococcus pneu
15	1306	57.4	1572	22	AAH90724	CFE 28 coding sequ
16	1306	57.4	7577	19	AAx52179	Streptococcus pneu
17	1305	57.4	792	20	AAx27222	S. aureus fff codi
18	1297	57.0	1569	20	AAx19484	Streptococcus pneu
19	1297	57.0	1569	24	ABK48379	DNA encoding strep
20	1279.5	56.3	1563	24	ABN67088	Streptococcus poly
21	1273.5	56.0	1563	24	ABN67087	Streptococcus poly
22	1266.5	55.7	2026	19	AAx296317	S. pneumoniae deri
23	1266.5	55.7	2026	19	AAx42992	Streptococcus pneu
24	1168	51.4	1641	22	AAH67220	C glutamicum codin
25	1168	51.4	349980	22	AAH68531	C glutamicum codin
26	1164.5	51.2	1362	22	AAH84550	E. coli growth and
27	1164.5	51.2	1362	23	AAx52529	E. coli DNA for ce
28	1158	50.9	721	23	AAx49372	Staphylococcus aur
29	1150	50.6	1374	23	AAx54213	Pseudomonas aerugi
30	1138	50.0	1389	23	AAx53231	Haemophilus influe
31	1138	50.0	1830121	17	AAx42063	Haemophilus influe
32	1133.5	49.8	1506	23	AAx56354	Salmonella typhi D
33	1131	49.7	8367	21	AAx81483	N. meningitidis pa
34	1131	49.7	349980	21	AAx21544	Neisseria meningit
35	1131	49.7	349980	21	AAx21607	Neisseria meningit
36	1115	49.0	92407	22	AAx28549	Genomic fragment #
37	1014	44.6	21567	23	AAx59514	Propionibacterium
38	1007	44.3	644	18	AAx75400	Staphylococcus aur
39	1006	44.2	640681	24	ABA92787	Buchnera sp. genom
40	907	39.9	910715	20	AAx20248	Borrelia burgdorfe
41	890	39.1	580073	18	AAx58840	Mycoplasma genital
42	854	37.6	1347	23	AAx53828	Helicobacter pylor
43	853	37.5	4804	20	AAx65141	Enrichia sp. E/4
44	847	37.2	521	23	AAx49408	Staphylococcus aur
45	847	37.2	521	23	AAx49436	Staphylococcus aur

ALIGNMENTS

RESULT 1  
AAx27221  
ID AAx27221 standard; DNA; 1368 BP.  
XX AAx27221;  
AC AAx27221;  
XX  
DT 28-MAY-1999 (first entry)  
XX  
DE S. aureus fff coding sequence.  
XX  
KW Ffh gene; signal recognition particle; SRP; antimicrobial agent; vaccine;  
KW immunological response; gene therapy; infection; otitis media;  
KW conjunctivitis; toxic shock syndrome; septic arthritis; ss.  
XX  
OS Staphylococcus aureus.  
XX  
PN EF902087-A2.  
XX

PD 17-MAR-1999.  
 XX PF 24-AUG-1998; 98EP-0306741.  
 XX PR 10-SEP-1997; 97US-0927216.  
 XX PA (SMIK ) SMITHKLINE BEECHAM.  
 XX PI Wallis NG;  
 XX WPI: 1999-169238/15.  
 DR P-PSDB; AAY00910.  
 DR XX  
 PT New Staphylococcus aureus Signal Recognition Particle (SRP) with  
 PT protein (fifh) and rna (fifh) components - the SRP gene and protein  
 PT useful as diagnostic reagents and for prevention and treatment of  
 PT Staphylococci infections which cause otitis media, septic arthritis  
 PT and toxic shock syndrome  
 XX PS  
 PS Claim 1; Page 28; 35pp; English.  
 XX CC  
 CC This sequence encodes the staphylococcus aureus signal recognition  
 CC particle (SRP) fifh component. Ffh polynucleotides are useful for  
 CC diagnosing a disease related to expression of ffh polypeptides by  
 CC analysing for the presence/amount of ffh protein in a sample due to  
 CC infection of a micro-organism with the gene, or determining the nucleic  
 CC acid sequence encoding ffh. Ffh polypeptides and antagonists are useful  
 CC for treatment of an individual in need (polypeptide) of, or needing to  
 CC inhibit (antagonist) ffh polypeptide levels. Ffh polypeptides and  
 CC polynucleotides are useful for identifying agonists and antagonists by  
 CC binding and observing the affect of ffh polypeptide activity, which are  
 CC potential anti-microbial agents. Ffh polypeptides and antigenic fragments  
 CC are also useful for inducing an immunological response (T cell/antibody)  
 CC to protect against disease, by direct administration (vaccine), or via a  
 CC vector (gene therapy). Anti-ffh antibodies are useful as antagonists, and  
 CC for protecting against disease. Diseases diagnosed, prevented and treated  
 CC include those caused by infection, especially bacterial infection,  
 CC including otitis media, conjunctivitis, toxic shock syndrome, wound  
 CC infection and septic arthritis. Ffh polypeptides are antimicrobial, and  
 CC are useful for bathing wounds and implants prior to surgical  
 CC implantation.  
 XX SQ

Sequence 1368 BP; 521 A; 186 C; 288 G; 373 T; 0 other;

#### Alignment Scores:

Pred. No.:	3 47e-197	Length:	1368
Score:	2274.00	Matches:	455
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	20	Gaps:	0

US-09-943-108a-2 (1-455) x AAX27221 (1-1368)

QY	1	MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys	20
Db	1	ATFGCAATTGAAGGCTATCAGACGCGCTGCAAGCGAGTGCACAAAGATGGTGGTAAAG	60
QY	21	GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
Db	61	GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAAGATTACGCTATT	120
QY	41	GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla	60
Db	121	GAGGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTTAAACACATATCAGACGGGCA	180
QY	61	LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln	80
Db	181	TTAGGTTCCGATGTAAATGCAATCATTAAACACCGGCAACAAAGTTATTAAATAGTTCAA	240
QY	81	AspGluLeuThrLysLeuMetGlyGluLysThrSerIleAsnMetSerAsnLysPro	100
Db	241	GATGAATTAACGAAGTTGATGGTGGAGAAAAATACATCGATTATATATGTCAAATAACCA	300
QY	101	ProThrValValMetValGlyLeuGlnGlyValAlaGlyLysThrThrAlaGlyLys	120
Db	301	CCTACTGTGTGTAIGAAGTGTGTTTACAAGTGCTGGTAAACAAACAACATCGAGTAAA	360
QY	121	LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle	140
Db	361	TTAGCATTATTGATGCGTAAAAAATACAAACAAAAACCTATGTTAGTTCACAGCATATT	420
QY	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
Db	421	TATGTCOCAGCAGCATAAATCAATTAACAAACAGTAGGAAACAAATTCATATCTCTGTA	480
QY	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
Db	481	TACAGTGAAGAGATGATGATGAAGCCACAAATTTGTAATCAATCAATTAACCATGCT	540
QY	181	LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu	200
Db	541	AAAGAAGAACATTTAGACTTTGTATATCATGTATACACAGCGTCGATTACACATCGATGAA	600
QY	201	AlaLeuMetAsnGluLysGluValLysGluIleAlaLysProAsnGluIleMetLeu	220
Db	601	GCATTGATGACGAATTAAGAGATGTAAGAAATTCCTAAACCAACGAAATTTATGTTA	660
QY	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln	240
Db	661	GTTGTCGATCAGTACAGGGTCAAGATGCTGCAATGTTCCAGAAATCTTTGACGATCAA	720
QY	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla	260
Db	721	CYTGATGTCACAGGTGTACCTTAACATAATAGATGCTGATACACAGTGGTGGTGCAGCT	780
QY	261	LeuSerIleArgSerValThrLysProIleLysPheValGlyMetSerGluLysLeu	280
Db	781	TTATCTATTCTGCTGGTGCACAAAAACCAATTAATTTGTTGTTGATGAGTGAAGGTTA	840
QY	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
Db	841	GATGTTTTAGAGCTATTTCATCTCCTGAACGTATGCGATCAGCTATTTTAGGTATGGGTAT	900
QY	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu	320
Db	901	GTGTTAAGTTTAAATGAAAAAGCGCAACAGATGCGATCAAGAAAAACCAAGATTTA	960
QY	321	GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln	340
Db	961	GAGAAAAAGATGCGTGAGTCATCGTTTACTTTTAGATGATTTTTAGAACACACTTCATCG	1020
QY	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
Db	1021	GTGAAAAATCTAGGACCATGGATGATATTATGAAAATGATTCACAGTATGATGAAAATG	1080
QY	361	LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle	380
Db	1081	AAAGGCGCTAGATAAGCTTAATATAGTGAAAGCAAAATTCATCATATTAAAGCGATTATC	1140
QY	381	GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys	400
Db	1141	CAGTCAATGACGGCGCTGAAAGAAACCAATCCAGACACATTAATGATGATCAGTAAAAAG	1200
QY	401	AspGluAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe	420
Db	1201	CGTATTCGTAAGGGTCTGGTGGTTCATTACAAGAGTCATTCGTTGATGAACATATT	1260
QY	421	AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys	440
Db	1261	AACGATATGAAGAAATGATGAACAAATTCCTGCTGGCGGTAAAGGTAAAAAGGTAAA	1320
QY	441	ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe	455
Db	1321	CGCAATCAATGCAAAATATGTTAAAGGTATGAAATTTACCGTTT	1365

## RESULT 2

AA555331  
ID AA555331 standard; DNA; 1368 BP.

XX AC AA555331;

XX DT 13-FEB-2002 (first entry)

XX DE Staphylococcus aureus DNA for cellular proliferation protein #1643.

XX KW Antisense; ds; prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX DR WPI: 2001-611495/70.

XX DR P-PSDB; RAU37472.

XX PT New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids -

XX PS Claim 27; Seq ID No 8968; 51pp; English.

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

XX ftp.wipo.int/pub/published\_pct\_sequences.

XX Seq Sequence 1368 BP; 517 A; 187 C; 291 G; 373 T; 0 other;

## Alignment Scores:

Pred. No.:	5,26e-196	Length:	1368
Score:	2261.00	Matches:	452
Percent Similarity:	99.78%	Conservative:	2
Best Local Similarity:	99.34%	Mismatches:	1
Query Match:	99.43%	Indels:	0
DB:	23	Gaps:	0

US-09-943-108a-2 (1-455) x AA555331 (1-1368)

QY	1	MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys	20
DB	1	ATGCAATTTGAGGGTTATCAGAACGCTTCAAGCGACGATGCAAAAAATGCGTGTAAAG	60
QY	21	GlyLysLeuThrGluAlaAspIleLysIleMetMetGluValArgLeuAlaLeuPhe	40
DB	61	GGTAAATCTACTGACGCTGATATAAGATATATGCGTGAAGTAAGATTAGCGTTACTT	120
QY	41	GluAlaAspValAsnPhelysValLysGluPheIleLysThrValSerGluArgAla	60
DB	121	CAGCGCTGACGTAACTTTAAAGTGTAAAGAAATTTATTAACACAGTATCAGACGCGCA	180
QY	61	LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln	80
DB	181	TTAGGTTCCGATGTAATGCAATCATTAACACAGCGCAACAAGTTATTAATAATAGTCAA	240
QY	81	AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
DB	241	GATCAATTAACGCACTGATGGGTGGAGAAATACGTCGATTAAATGTCGAATAACCA	300
QY	101	ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys	120
DB	301	CCTACTCTTCTGTTATGATGGTTGGTTTACAAGGTCCTGGTAAACACAACTGCAGGTAA	360
QY	121	LeuAlaLeuLeuMetArgLysLysTyraAsnLysLysProMetLeuValAlaAlaAspIle	140
DB	361	TTAGCATTTGATGGTAAATAATACACAAACCACTTGTAGTTCCAGCAGATATT	420
QY	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
DB	421	TATGTCACGACGATAAATCAATTAACACAGTAGGGAACAAATGATATCTCTGTA	480
QY	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
DB	481	TACAGTGAAGGAGATCAAGTAAAGCCACAATAATGTAACTAATGCAATGCAATGCAATGCT	540
QY	181	LysGluGluHisLeuAspPheValIleLeuThrAlaGlyArgLeuHisIleAspGlu	200
DB	541	AAAGAAACAACTTTAGACTTTGTAATCATTCATACAGCAGGTGCGATTACATCGATGAA	600
QY	201	AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu	220
DB	601	CGATTGATGAACCAATTAAGAGTAAAGACATGCTTAAACCAACCAAAATATGTTA	660
QY	221	ValValAspSerMetThrGlyClnAspAlaValAsnValAlaGluSerPheAspAspGln	240
DB	661	GTGTGCGATTCAATGACGGGTCAAGTGTCTCAATGTTGCGAGATCTTTTGACGATCAA	720
QY	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyClyAlaAla	260
DB	721	CTTGATGTCACAGGTGTTACCTTAATAATAGATGGTGATACACGTGGTGGCGAGCT	780
QY	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
DB	781	TTATCTATTGCTTCGCTGACAAAAACCAATTAATTTTGGTATGAGTCAAAAGTTA	840
QY	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
DB	841	GATGGTTTAGAGCTATTCCTCTGACGATGATGGCATCAGCTATTTTAGGTATGGGTGAT	900
QY	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu	320
DB	901	GTGTTAAGTTTAATTCGAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAAAGATTTA	960
QY	321	GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln	340
DB	961	CAGAAAAAGATGCGTGCATCGTTTACTTTAGATGATTTTGTAGAACAACTTGCATCAG	1020
QY	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
DB	1021	GTGAAAAATCTAGGACCACTGGATGATTAATGAAAAATGATTCAGGATGAATAAATG	1080



QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
DB 781 TTATCTATTGGTTGGTGACAAAACCAATTAAATTTGGTTGGTATGAGTGAAGATTA 840  
QY 281 AspGlyLeuLysPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
DB 841 GATGGTTAGAGCTATCCATCGTAGGATGGCATCAGCTATTCAGGCGCTGGGTGAT 900  
QY 301 ValLeuSerIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
DB 901 GTGTCAAGGTTAATGAAAAAGCGACACAGATGTGATCAAGAAAAAGCAAGATTTA 960  
QY 321 GluLysLysMetArgGluSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340  
DB 961 GAGAAAAAGATGGCGAGTCATCTTTTATGATGATTTTATGACAACTTGTATCAG 1020  
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
DB 1021 GTGAAAAATTTAGACCACTGGATGATATTATTAATGATTCAGGTATGAATAAATG 1080  
QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnLysHisIleLysAlaIleLe 380  
DB 1081 AAAGCTAGATAAGCTTAATATGATGAAAAAGCAAAATGATCATATTAAAGCGATTATC 1140  
QY 381 GlnSerMetThrProLagluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
DB 1141 CAGTCATGACGGCGGTGAAGAAACAAATCCAGACACATGATGATCAGCTAAAGAG 1200  
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
DB 1201 CGTATTGCTAAAGGTTCTGGTGTTCATTACAGAAAGTCAATCGTTGATGAACAATTT 1260  
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysLys 440  
DB 1261 AACGATATGAGAAATATGATGAACAACTCACTGGTGGCGGTAAAGGTAAAGGTAAA 1320  
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
DB 1321 CGCAATCAATGCATAATATGTATAAGGTATGATTTACCGTTT 1365

RESULT 4  
AAS54393  
ID AAS54393 standard; DNA; 1368 BP.  
XX  
AC AAS54393;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Staphylococcus aureus DNA for cellular proliferation protein #705.  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO200170955-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.  
DR P-PSDB; AAU36534.  
XX  
PT New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Claim 27; Seq ID No 8030; 511pp; English.  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1368 BP; 520 A; 190 C; 285 G; 373 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 1-21e-193 Length: 1368  
Score: 2235.00 Matches: 447  
Percent Similarity: 98.90% Conservative: 3  
Best Local Similarity: 98.24% Mismatches: 5  
Query Match: 98.28% Indels: 0  
DB: 23 Gaps: 0  
  
US-09-943-108A-2 (1-455) x AAS54393 (1-1368)  
  
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
DB 1 ATGGCATTTGAAGGCTTATCAGAACGCTGCAAGCAGCATGCAAAAAATGGTGGTAAAG 60  
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgLysValArgLeuAlaLeuPhe 40  
DB 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAAGATTAGCGTTACTT 120  
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
DB 121 GAGGCTGACGTAACTTAAAGTGTAAAGAAATTTATTAACACATATCAGACGGCA 180  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80  
DB 181 TTAGTTCGGATGTAATGCAATCAATTAACACAGGCAACAACTATTTAAATAGTTCAA 240  
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
DB 241 GATGAATTAACGAGTGTGATGGGTGGAGAAAATATATCGATTAAATATGCAAAATAAGCA 300  
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
DB 301 CTTACTGTGTATGATGTTGGTTTACAGGTGCTGGTAAACACACACACACAGTAA 360  
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
DB 361 TTAGCATTTATGATGCGTAAATAATACAAACAAAAAACCCTATGTTAGTTCCAGCATATT 420  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
DB 421 TATCGTCCAGCAGCATTAATCAATTAACAAACAGTAGGGAACAAATTCATTCCTGTA 480

```
QY 161 TyrSerGluGlyAspGlnValLysProGlnInleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGAGATCAAGTAAAGCCACAAATTTGTAATCATGATTAACATGCT 540
QY 181 LysGluGluHisLeuAspPheValIleLeaspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACAATTTAGACTTTGTATCATATGATACAGCAGGTGATTAACATCATGATCAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCRITGATGAATGAATTAAGAAGTAAGAAGACATTTCTTAACCAACCAATTAATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATTCATTAATGACGGTCAAGATCTGTCAATGTTGCAGAAATCTTTTGACGATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGATGTACAGGTGTTACCTTAATTAATAGATGGTGATACAGCTGGTGGTCAGCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTAATCTATTCCTGCGACACAAACCAATTAATTTGTTGATGAGTGAARAGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGCTTTAGAGCTATTCCATCTCAAGCTATGCGATCACTGATTTTCAGGCGCTGGTGTAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTCAAGGTTAAITGAAAAGCGCAACAAGATGTGGATCAAGAAAACCAAAAGATTTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAGATGCGCGAGTCAATCTTACTTTAGTGAATTTTAAAGCAACATTCATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATTTAGACCACTGGATGATATTATTAATGATTCAGGTATGAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGTCTAGATAAGCTTAATATGATGAAGCAAAATGATCATATTAACCGGATTATC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAsnProAsnThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACCGCGCTGAAAGAAACATCCAGACACATTTGATGTATCAGTAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTAAAGGTTCTGTCCTTCATTACAAGAAGTCAATCGTTTGATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysGlyLysGlyLys 440
Db 1261 AACGATAGAGAAGAAATGTAACCAATTTCACTGGTGGCGGTAAAGTAAAGGTAAAG 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACGGITT 1365
RESULT 5
ABQ70511/c
ID ABQ70511 standard; DNA; 2256 BP.
XX
AC ABQ70511;
XX
XX 29-AUG-2002 (first entry)
XX
DT Listeria monocytogenes 4b contig DNA sequence #453.
XX
DE Antibacterial; Listeria; food contamination; mutational analysis;
XX
KW infection; ds.
XX
```

```
OS Listeria monocytogenes 4b.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
XX (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Kunst F, Glaser P;
XX
XX WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX
PS Claim 14; SEQ ID 3324; 180pp; French.
XX
XX The present invention relates to nucleic acid sequences
CC (AB067188-AB0711212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2256 BP; 533 A; 471 C; 420 G; 832 T; 0 other;

Alignment Scores:
Pred. No.: 1-45e-146 Length: 2256
Score: 1719.50 Matches: 331
Percent Similarity: 86.81% Conservative: 64
Best local Similarity: 72.75% Mismatches: 55
Query Match: 75.62% Indels: 5
DB: 24 Gaps: 1

US-09-943-108A-2 (1-455) x ABQ70511 (1-2256)
QY 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1666 ATGGCATTTGAAGGACTAGCTGGAGACTCCCAAGAAACAATGAACAAATTCGCGGCAAA 1607
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 1606 GGAAGAAGTAACAGAGCTGACGTAAAGAAGAAATGATCGGTGAAGTTCGTCTGCTACTT 1547
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 1546 GAAGCGGATGTTAACTTTAAAGTCGTTAAACAATTTATTAACAACAGTAAGCGAAGTCT 1487
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 1486 GTGCGCGGAGCTTATGAAGAAGCCTTAACACCGCGCAACAAGTTATCAAAATTCGTTCAA 1427
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 1426 GAAGAAGTAACAGCTTAATGGCGGAGGAAGAAAGCAAAATTCGGAACACGCGACCGCCA 1367
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 1366 CCACCGGTATTAATGATGAGTTTACAGGGGGCTGGTAAACACAGGACTTCCGCGCAA 1307
QY 121 LeuAlaLeuMetMetArgLysLysTyrAsnLysLysProMetLeuValAlaAspIle 140
```





```
Db 188164 GAAGAATTACAAGCCTCAAGGGCGGCGAAGAAAGTAAATCGGAACAGCGATCGCCCG 188223
QY 101 ProThrValValMetValGlyLeuGlnGlyValGlyLeuGlyThrThrAlaGlyLys 120
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188224 CCACCGTTATTATGAGTGTAGTTTACAGAGGCTGTGTAAACACACCTTCAGGAAA 188283
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188284 CTCGCTAATTATTACGTAAATAATATCGTAAACCTTTACTAGTCGCGAGCAGATATT 188343
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188344 TACCGACCCGACGAATCAACCAATAGAAACACTGTGCAACCAATAGATATGCGCGTA 188403
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188404 TTTTCTCAGGGCATCAAGTAAGCCCACTAGAAATCGCGAACAAGCTATCGCTAAAGCA 188463
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188464 AAAGAAGAACATTTAGATTATGTCTATTCGATACAGCTGCTCTCATATCGACGAA 188523
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188524 ACTCTGATGGAGCAATTAACAAAGTGAAGAAATCGTACGCCAACTGAAATTTTACTT 188583
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188584 GTAGTTGATTCAATGACTGGCGACGACGAGTAAATGTGTGCGCCAAAGCTTCACGACAA 188643
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188644 TTAGAAATTTACCGCGTGTATTAAACAAATTAGACGCTGATACACGTGTGGGGCAGCA 188703
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188704 CTTTCATCCGTCAGTCACAGGAAACCAATCAAAATTCGTGTACCCGGTGAAGAAATG 188763
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188764 GAAGCAATCGAAACGTTCCATCCGATCGTATGGCTTCAAGAAATTCGCGCATGGGTAT 188823
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGlnLysAlaLysAspLeu 320
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188824 GTACTTCTCTATTGAAAAAGCACAACTGATGATGACAGAAAAAATGAAAGCTATG 188883
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnIleLeuAspGln 340
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188884 GACAAAATGAAGAACACACAGCATGACCTTAGATGACTTCTTGGACCAATGCAACAA 188943
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188944 GTAAACAAATGGGACCACTAGATGAATCTACTATAAATGATGCCAGGGCAACAAATG 189003
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLeu 380
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189004 AAAGGCGCTCGAACATGAATGATGATAAACAACACTCGGTGCACATCGAAGCGATAAT 189063
QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189064 AAATCCATGACCAAAACAGAAAGATAATATCCGACATCATCAATGCGACAGAGAGAAA 189123
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189124 CGAATGCTCGTGGAGGAGCGGCCCAATTCAGAAATCAATCCCTCCCTTAACAATTT 189183
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189184 GCTGAAATGAAAAAATGATGAAGCAAAATGACTGCTGGAGGAGAAAGTGAAGAGGTAA 189243
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189244, -----AATCCATTCGGCAATTTTCAAAATGCCATTT 189273
```

```
RESULT 7
ABN92495
ID ABN92495 standard; DNA; 1014 BP.
XX
AC ABN92495;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1958.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX
KW antibacterial; gene therapy; gene; ds.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
(GENO-) GENOME THERAPEUTICS CORP.
XX
Doucette-Stamm LA, Bush D;
XX
WPI; 2002-381255/41.
DR P-PSDB; ABP39950.
XX
Novel isolated nucleic acid encoding a Staphylococcus epidermidis
polypeptide, useful for diagnosing and treating bacterial infections -
XX
Disclosure; SEQ ID 1958; 267pp; English.
XX
ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
frame (ORF) nucleic acid sequences which encode the amino acid sequences
given in ABP35124 to ABP37960. The S. epidermidis sequences have
antibacterial activity and can be used in gene therapy. The sequences
can also be used in the diagnosis and treatment of bacterial infections,
particularly S. epidermidis infections. The sequences can be used to
screen for compounds able to interfere with the S. epidermidis life
cycle or inhibit S. epidermidis infection.
XX
N.B.: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from the
USPTO web site.
XX
SQ Sequence 1014 BP; 354 A; 156 C; 221 G; 283 T; 0 other;
XX
Alignment Scores:
Pred. No.: 3.75e-130 Length: 1014
Score: 1534.00 Matches: 307
Percent Similarity: 97.29% Conservative: 16
Best Local Similarity: 92.47% Mismatches: 9
Query Match: 67.46% Indels: 0
DB: 24 Gaps: 0
US-09-943-108A-2 (1-455) x ABN92495 (1-1014)
QY 1 MetaAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 ATGCGATTGTGAGGATTTATTCGATCGCTTACAGCCACGATGCAAAAATGCGTGTATA 75
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 GGAAGAAGTAACAGAAGCAGATATTAAACAGATGATGCGTGAAGTGAGATTAGCTATTG 135
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluLysAla 60
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 GAAGCGGATGTTAACTTCAAAAGTTGTTAAGGAATTTGTTAAGAAATGTTTCAGAACGAGC 195
```

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGln 80  
 DB 196 CTAGGTTCTGATGTCAGCAATCTTTAACACCTGGCCACAGGTTATTAATATCGTACAA 255  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 DB 256 GAGACACTACTAGTTTAAATGGGAGGAAATACCTCCATTAGATGGCAACAACCA 315  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 DB 316 CCAACTGTGTCTCATGTCGCTTACAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 375  
 QY 121 LeuAlaLeuLeuMetAspLysLysTyrAsnLysLysProMetLeuValAlaAspPhe 140  
 DB 376 TTGGCATTTAATCGTAAATAATATACAAATACCTTACTGTGGCAGAGATAT 435  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 DB 436 TATCGTCCAGCTGCTATTGATCAATTACAAACAGTAGTAAACAAATGATATCCCTGTG 495  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 DB 496 TATAGTGAAGGTGATCAAGTATATCAACACAAATGTTGAAATGCTTTAAACATGCT 555  
 QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
 DB 556 AAGAGACACATCTAGATTCCTTAATCATGATGATACAGCTGCTTTACACATGTGAA 615  
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 DB 616 GCATTTATGATGAGCTTCAAGAGTAAAGAAATCTCTAAACACAGCAAAATATGCTT 675  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
 DB 676 GTTGTGTGATGATGACAGGTCAGATGCTGTGAATGTGGCACAATCATTTGATGACAA 735  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260  
 DB 736 TTAGATGTTTCAGGTGTAACATTGACTAAATAGATGATGATACACGCGTGTGACGA 795  
 QY 261 LeuSerIleArgSerValThrGluLysProIleLysPheValGlyMetSerGluLysLeu 280  
 DB 796 CTTTCAATCCGTTCCGTTACCCAAAACCTATTAAATTTGTAGGTATGATGAGAAATG 855  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 DB 856 GATGGTTTGAATATTATTCCTCCTGAACGAATGCCCTCCTATTTTAGGTATGGGTAT 915  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 DB 916 GTTTTAAAGCTCATTTGAAAAGCGCAACAGATGTAGATCAAGAAAAGGCTAAAGATTTA 975  
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAsp 332  
 DB 976 GAGAAGAAGATGAGAGATCATCTCTTCTCTAAAT 1011  
 RESULT 8  
 AAS51308  
 ID AAS51308 standard; DNA; 1416 BP.  
 XX  
 AC AAS51308;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Enterococcus faecalis DNA for cellular proliferation protein #85.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 FN WO200170955-A2.  
 XX

PD 27-SEP-2001.  
 XX  
 EF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2001-611495/70.  
 DR P-PSDB; AAU33449.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 FS Claim 27; Seq ID No 3890; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1416 BP; 494 A; 228 C; 337 G; 357 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.74e-119 Length: 1416  
 Score: 1418.50 Matches: 273  
 Percent Similarity: 77.54% Conservative: 86  
 Best Local Similarity: 58.96% Mismatches: 89  
 Query Match: 62.38% Indels: 15  
 DB: 23 Gaps: 2  
 US-09-943-108A-2 (1-455) x AAS51308 (1-1416)  
 QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 DB 1 ATGGCTTTTGAGAGTTTACAAACCGCTACACAGGCAATGAGTAAATCCGTCGTAG 60  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 DB 61 GGAAAAATTTCCGAAGCGAGCTAAAGAAATGATCGAGAAATCCGTTGGCTTTATTA 120  
 QY 41 GluAlaAspValAsnPhelysValIlyLysGluPheIleLysThrValSerGluArgAla 60  
 DB 121 GAAGCCGACGTTAAATTTCAAGTGTGTCACAAAGATTTCACAAAACCGCTCAGACAGCGGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 DB 181 GTAGGATCGAAGTATTAGAAGCTTATCACCAGCCCAAAATTTGTAATAATTTGTTGAT 240

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
DB 241 GAAGAAATCAAGAAACCTGTAGGTGCAAGACGGTTGAACCTGAATCAATATCTCCAAATC 300  
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
DB 301 CCGACGTGATATGATGACAGGGTTACAGGGGCTGGTAAACACACTTTTACTGGTAA 360  
QY 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspIle 140  
DB 361 TTACCAAAACACCTTAATGAACACGTAAACACGTCGTCGCTTTAATCGCTGGTGAGTT 420  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
DB 421 TATCGTCCACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180  
DB 481 TTTGATATGGAAACAGATGCTAATCCAGTGGAAATTTCTGCAAGGTTTAGCATAGCA 540  
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
DB 541 AAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
QY 201 AlaLeuMetAsnGlnLeuLysGlyValLysGlnIleAlaLysProAsnGlnIleMetLeu 220  
DB 601 GCTTATATGACCAATGAAACAAATTAAGAGTTTGGCTAATCCCAATGAATTTCTGTTA 660  
QY 221 ValValAspSerMetThrGlyAlaAspAlaValAsnValAlaGluSerPheAspAspGln 240  
DB 661 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260  
DB 721 CTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780  
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
DB 781 CTGTCAATTCGGCAGTAAACGGCGCTCCGATTAATTTGTCGGTTCGTTGGAATTA 840  
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
DB 841 ACCGATTTAGAAATTTCCATCCCGCATGATGTCAGTGCATGCTAGTATGATGATGATG 900  
QY 301 ValLeuSerIleLeuLysAlaGlnAlaAspValAspGlnGlnLysAlaLysAspLeu 320  
DB 901 ATGTTGACGCTAATGAAAGACCCACACAGATTACGATGAGAAAGACAGAGAACTT 960  
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340  
DB 961 GCTCAAAATGAAAGAAACAGTTTTCACCTTAAACGATTTCAATGACCAATGGATCAA 1020  
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
DB 1021 GTTATGGGATGGGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
DB 1081 CCTGGTATGAAATGTCAAAGTGCATCAAAAGATGTGGCAGCAAGACGGCGATGTC 1140  
QY 381 GlnSerMetThrProAlaGluArgAsnAspProAspThrLeuAsnValSerArgLysLys 400  
DB 1141 CTATCATGACCCCTGCAAGACCGTAAATCCCTGATCTATTAACTTACTGTCGCGTGC 1200  
QY 401 ArgIleAlaLysLysSerGlyArgSerLysGlnGluValAspArgLeuMetLysGlnPhe 420  
DB 1201 AGAATTCACCTGGTTCAGGAAATAGTGTGTTGAAGTCAATCGTATGATTAACAAATT 1260  
QY 421 AsnAspMetLysMetLysMetLysGlnPheThr----- 431  
DB 1261 AAAGAAATCCAAAGAAATGATGCAACAAATGTCCAAGGGGATATCAACATTCCTGGTATG 1320  
QY 432 -----GlyGlyGlyLysGlyLysArgAsnGlnMetGlnAsn 446

DB 1321 GATCAAAATGCTAGGTGGCGGCTTAAAGCAAGTAGTAAATG---GCCATGAATCGT 1377  
QY 447 MetLeuLys 449  
DB 1378 ATGATGAAG 1386  
RESULT 9  
AAS53127  
ID AAS53127 standard; DNA; 1434 BP.  
XX  
AC AAS53127;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Enterococcus faecalis DNA for cellular proliferation protein #555.  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Enterococcus faecalis.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
DR WPI; 2001-611495/70.  
DR P-PSDB; AAU35268.  
XX  
XX New polynucleotides for the identification and development of  
antibiotics, comprise sequences of antisense nucleic acids -  
PS Claim 27; Seq ID No 6764; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
prokaryotic cellular proliferation, their use in identifying the  
genes, their use in the discovery of novel antibiotics, the essential  
genes themselves and the encoded proteins. The prokaryotes used are  
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
invention is also useful for the identification of potential new targets  
for antibiotic development. The antisense nucleic acids can also be used  
to identify proteins used in proliferation, to express these proteins,  
and to obtain antibodies capable of binding to the expressed proteins.  
XX  
XX The proteins can be used to screen compounds in rational drug discovery  
programmes. The antisense nucleic acid sequence is also useful to screen  
for homologous nucleic acids which are required for cell proliferation in  
a wide variety of organisms. The present sequence encodes an  
essential prokaryotic cellular proliferation protein.  
XX  
XX Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
cc ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1434 BP; 500 A; 230 C; 340 G; 364 T; 0 other;  
Alignment Scores: 1.77e-119 Length: 1434  
Pred. No.:

Score: 1418.50 Matches: 273  
Percent Similarity: 77.94% Conservative: 86  
Best Local Similarity: 58.96% Indels: 89  
Query Match: 62.38% Gaps: 15  
DB: 23

US-09-943-108a-2 (1-455) x RAS53127 (1-1434)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
Db 1 ATGGCTTTTTCAGAGTTTAAACAACCGCTACACAGGCAATCAGTAAATCCGCTGTAAG 60  
QY 21 GlyLysLeuThrCluAlaAspLleIysLleMetMetArgGluValArgLeuAlaLeuPhe 40  
Db 61 GGAAGAATTTCGAGAGCGAGCTAAAGAAATGTCGGAATCCGTTGGCTTATTA 120  
QY 41 GluAlaAspValAsnPhelValValLysGluPheLleLysThrValSerGluArgAla 60  
Db 121 GAAGCCGAGCTTAATTTACAAGTGTCAAAAGATTTCACAAAACGGCTCAGAGAACGGCA 180  
QY 61 LeuGlySerAspValMetClnSerLeuThrProGlyGlnGlnValLleLysLleValGln 80  
Db 181 GTAGAGTCGAGATATTGAAGCTTATCCACGACCCCAAAATGTGAATTTGTTGAT 240  
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
Db 241 GAAGAATTACGAAACGCTAGGTCAGAAACGGTTGAATGAATTAATCTCCAAAATC 300  
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
Db 301 CCGACAGTGATTATGATCAGAGGTTTCAAGGGCTGTAAGAAACAACTTTTACTGGTAAA 360  
QY 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspLle 140  
Db 361 TTACGAAACACATTAATGAACACTGAACACGCTGCGCCCTTTTAATCGCTGGTGAGTT 420  
QY 141 TyrArgProAlaAlaLleAsnGlnLeuGlnThrValGlyLysGlnLleAspLleProVal 160  
Db 421 TATCGTCCAGCAGCGATTGATCAGTTGAGGTTTAGGTCACAAATTAAGTTCCTCCGTT 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLleValThrAsnAlaLeuLysHisAla 180  
Db 481 TTTGATATGGGACACAGATCTAATCCAGTGGAAATTTGTCGTCAGGGTTAGCATAGCA 540  
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
Db 541 AAAGAAAGAAATTAATTAATCTCTTAATGATACGCGCGCGCTTACACATTGACGAA 600  
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLleAlaLysProAsnGluLleMetLeu 220  
Db 601 GCTTTAATGACGAAATGAACAAATTAAGAGTTGGCTAATCCCAATGAATTCGTGA 660  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
Db 661 GTTGTGATCGGAGTACGGGGACAGATGCTGTCAAGTTGACAGATAGTTTAATGACAG 720  
QY 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
Db 721 CTGGAATTTACTGGGTTTGTATTACCAAAATGAACGCGGATCTCGTGGGGGGCTGG 780  
QY 261 LeuSerIleArgSerValThrClnLysProLleLysPheValGlyMetSerGluLysLeu 280  
Db 781 CTGTCAATTCGGGACATACGGCGCTCCGATTAAATTTGCGGTCGTGGTGAATAATTA 840  
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
Db 841 ACCGATTTAGAAATTTTCCATCCCGATGATGTCGAGTCGTATCCTAGGTATGGGGGAC 900  
QY 301 ValLeuSerLeuLleGlnLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLys 320  
Db 901 ATGTTGACGCTAATTTGAACACGCAACAGATTAAGTACGATCAGAAAAGCAGAGACACT 960  
QY 321 GluLysLysMetArgGluSerPheThrLeuAspPheLeuGlnLeuAspGln 340

Db 961 GCTCAAAAATGAAGAAACAGTTTTCAGTTTACAGTTTTCATTTGAGCAATGGATCAA 1020  
QY 341 ValLysAsnLeuGlyProLeuAspAspLleMetLysMetLleProGlyMetAsnLysMet 360  
Db 1021 GTTATGGCATGGGACCGATTGAAGACTTATTAAATATGATCCTCGTGAATGATCAATG 1080  
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnLleAspHisLleLysAlaLle 380  
Db 1081 CTGGTATTGAAATGTCAAAATCGATCCAAAGATGTGGCAGGAACGGCGATGTC 1140  
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
Db 1141 CTATCATGACCCCTCGACAGACGTGAATCTGATCTATTAAATCTAGTCGCCGTGCG 1200  
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
Db 1201 AGAATTGCACTGGTTCAGGAATAGTGTGTTGAAGTCAATCTATGATTAACAATTT 1260  
QY 421 AsnAspMetLysLysMetMetLysGlnPheThr 431  
Db 1261 AAAGAAATCCAAAATAATGTCACACAAATGTCCTCAAGGGGATATGAACATTCCTGTTAG 1320  
QY 432 -----GlyGlyGlyLysLysGlyLysLysGlyLysArgAsnGlnMetGlnAsn 446  
Db 1321 GATCAAAATGCTAGTGGCGGCTTAAAGCAAGTTAGTAAATG---GCCATGAATCGT 1377  
QY 447 MetLeuLys 449  
Db 1378 ATGATGAAG 1386

RESULT 10  
AXL12974  
ID AXL12974 standard; DNA; 6729 BP.  
XX AAX12974;  
XX 19-MAR-1999 (first entry)  
XX Enterococcus faecalis genome contig SEQ ID NO:37.  
XX Enterococcus faecalis; contig; detection; Enterococcal infection;  
XX vaccine; attenuation; computer readable medium; ds.  
XX Enterococcus faecalis.  
XX WO980555-42.  
XX 12-NOV-1998.  
XX 04-MAY-1998; 98WO-US08985.  
XX 14-NOV-1997; 97US-0066009.  
XX 06-MAY-1997; 97US-0044031.  
XX 16-MAY-1997; 97US-0046655.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Barash SC, Dillon PJ, Kunsch CA;  
XX WPI; 1999-045171/04.  
XX New isolated Enterococcus faecalis polynucleotides and polypeptides  
XX - used to develop products for the detection of Enterococcus and for  
XX use in vaccines for prevention or attenuation of Enterococcus  
XX infection.  
XX Claim 1; Page 406-409; 2084pp; English.  
XX A computer readable medium has been developed which has recorded on it  
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
XX AXL12938 to AXL1919 represent these nucleotide sequences which are  
XX primary nucleotide sequences, also known as contigs. The computer-based

system can identify fragments of the *Enterococcus faecalis* genome with commercial importance. The products can be used to detect the presence of *Enterococcus faecalis* in samples. They can also be used for diagnosing *Enterococcal* infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, *in vivo* or *in vitro*. In particular the polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to prevent or attenuate an *Enterococcal* infection.



PR 30-DEC-1999; 99US-0174089.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;  
 PI Thanassi JA;  
 XX WPI; 2001-496721/54.  
 XX  
 XX Nucleic acids encoding conserved essential genes involved in bacterial  
 PT replication which are potential targets for the treatment of antibiotic  
 PT resistant bacterial infections -  
 XX  
 XX Claim 30; Fig 49; 380pp; English.  
 XX  
 CC The present invention relates to nucleic acids (AAH90701-AAH90918)  
 CC encoding polypeptides (AAH01002-AAH01114), which are essential for the  
 CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For  
 CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic  
 CC acids are useful for detecting the presence of proteins essential for the  
 CC viability of a bacterial cell wall in samples such as cells, tissues,  
 CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,  
 CC and for detecting corresponding target nucleic acid molecules with  
 CC complementary sequences. The nucleic acids are also useful for  
 CC determining whether a genomic nucleotide sequence of interest is  
 CC essential for viability of a bacterial cell or whether it resides within  
 CC an operon, by integrating an exogenous nucleotide sequence comprising a  
 CC portion of an open reading frame of the genomic sequence of interest  
 CC (comprising 200-500 base pairs) into the genomic sequence of interest  
 CC which confers a selectable phenotype to the cell, and determining cell  
 CC viability with a selection agent such as chloramphenicol. The nucleic  
 CC acids and proteins are also useful as vaccines and for treating bacterial  
 CC infections with gene therapy and antisense therapy. The nucleic acids  
 CC also enable identification of targets suitable for the treatment of  
 CC antibiotic resistant bacterial infections.  
 XX  
 SQ Sequence 1605 BP; 484 A; 317 C; 393 G; 411 T; 0 other;

## Alignment Scores:

Pred. No.: 9.54e-110 Length: 1605  
 Score: 1312.00 Matches: 254  
 Percent Similarity: 74.95% Conservative: 90  
 Best Local Similarity: 55.34% Mismatches: 101  
 Query Match: 57.70% Indels: 14  
 DB: 22 Gaps: 2

US-09-943-108A-2 (1-455) x AAH90837 (1-1605)

QY 1 MetalAPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 DB 1 ATGGCATTTGAAGTTTAAACAGAACGTTTGAGAACGCTTTTAAATAATCTACGTAATAAAA 60  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 DB 61 GGAATAATCTGTAATCTGATGTCAGAGGACCAACCAAGGAGTATGCTTGGCTTGGCTC 120  
 QY 41 GluAlaAspValAsnPhetLysValValLysGluPheIleLysThrValSerGluArgala 60  
 DB 121 GAGGCCGACGTTGCTTGCCTGTTGTAAGGACCTTATCAAGAAAGTATGAGCGTGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 DB 181 GTCGGGCATGAGGTCATGTGATCACTTAATCCTCGCACACAGATTATTAATAATCCTTGTAT 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 DB 241 GAGGACTGACACCGCTTTAGTTCTGTACGCGAGAAATATCAAGTCACCTAGATT 300  
 QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 DB 301 CCAACCATCATCATGATGTTGTTTACAGGGGCTGGTAAACACCTTTGCTGGTAA 360  
 QY 121. LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140

DB 361 TTGCCCAACAACTCAGAAAGAAAGAAATCTCTCTTGTATGCTTGGTCCGCGGATATT 420  
 QY 141 TyrArgProAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 DB 421 TATGTCACAGTCCATTTACACGCTTAAGACCTTGGGACACACAGATTGATGCTGCTGC 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 DB 481 TTTCACCTTGGACACAGAACTACACCTCTTGAGATTGTACGTCAAGGTTTGGACAGGCC 540  
 QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
 DB 541 CAACTAATCATCAACGACTATCTCTTGATTGATCTCGGGTCTGTTGCAGATTGATGAG 600  
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGlnIleAlaLysProAsnGlnIleMetLeu 220  
 DB 601 CTCCTCATGATGAGTCTGCTGATGTGAACATTTGGCTCAACCAATGAATCTCTT 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
 DB 661 GTCGTTGATGCTATGATTGTCAGGAACACCAATGTTCCGCGTGAGTTTAAATGCTCAG 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 DB 721 TTGGAAGTGAATGGGCTCATCTCTTACCAAGATTGATGGCGATACTCGTGGTGGTCTGCT 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 DB 781 CTGCTGTGTCACATTTACTGGAAACCAATCAAGTTCTGTTACAGGTGAAAGAAAT 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 DB 841 ACGGACATTTGAACCTTCCACCCAGACCGCATGCTAGCCGTATCTCTTGGTATGGGGAT 900  
 QY 301 ValLeuSerIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
 DB 901 ATGCTCACTTTGATTGAGAAAGCTTCTCAGGAATACGATGACAAAGAGCCCTTGAATG 960  
 QY 321 GlnLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340  
 DB 961 GCTGAGAGATGCGCGAAACACCTTTGATTTTAACTTATCATCATCAATAGATCAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 DB 1021 GTGCAAAATATGGGCGGATGGAAGACTTGTCTCAAGATGATTCAGGTATGCCCAACAAT 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
 DB 1081 CCAGCCCTTCAAAACATGAAGGTGATGAACGCCAGATTGCTCGTAACGTCCTCATTTGTG 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 DB 1141 TCTTCGATGACACCTGAGAGCGGTGAACACCCAGATTGTTTAATCCAGCGCTCGCGCT 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
 DB 1201 CGTATTGCTGCTGTTCTGAAATACATTCGCAAGTCAATAAATTCATCAAGGACTTT 1260  
 QY 421 AsnAspMetLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLys 440  
 DB 1261 AACCGGCTAAACAGCTCATGCGAGGCTGTTATGCTGGG-----AsnLeuPro 454  
 QY 441 ArgAsnGlnMetGlnMetLeuLysGlyMet-----GATAATGAATAAATGATGAGCAAAATGGGATTAATCCAAATAACCTTCT 1350  
 DB 1300 -----GATAATGAATAAATGATGAGCAAAATGGGATTAATCCAAATAACCTTCT 1350  
 RESULT 13  
 ID AAS55519  
 ID AAS55519 standard; DNA; 1572 BP.  
 XX  
 AC AAS55519;  
 XX



DT 13-FEB-2002 (first entry)  
 XX Streptococcus pneumoniae DNA for cellular proliferation protein #90.  
 DE Antisense: ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 OS Streptococcus pneumoniae.  
 XX  
 XX WO200170955-A2.  
 PN 27-SEP-2001.  
 XX  
 XX 21-MAR-2001; 2001WO-US09180.  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 DR P-PSDB; AAU37660.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Claim 27; Seq ID No 9156; 511pp; English.  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1572 BP; 476 A; 300 C; 386 G; 410 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 2.65e-109 Length: 1572  
 Score: 1307.00 Matches: 253  
 Percent Similarity: 74.73% Conservative: 90  
 Best Local Similarity: 55.12% Mismatches: 102  
 Query Match: 57.48% Indels: 14  
 DB: 23 Gaps: 2

US-09-943-108a-2 (1-455) x AAS55519 (1-1572)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGGATTGTAAGATTAAACAGACGTTTGCGAGAACGCTCTTTAAATAATCTACGTTAAAAA 60  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

Db 61 GGAATAATCTCTGAATCTGATCTCCAGAGCGCAACCAAGAAATTCGCTTGGCCTTGC 120  
 QY 41 GluAlaaspValasnPheLysValLysValLysGluPheLeuLysThrValSerGluArgAla 60  
 Db 121 GAGCGGACGCTTGCCTTGCCTTGAAGGACATTTATCAAGAAGTTTCGTGCGGTGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnValLysLeuValGln 80  
 Db 181 GTGCGGATGAGTGCATTGATACCTTAATCCTGCGCAACAGATTAATAATCGTGAT 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 241 GAGGAATGACACCCCTTTAGTTCTGATACGCGCAAAATTAACAAGTACCTAAGATT 300  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCAACCATCATCATGATGTTGTTTACAGGGCTGTGTAACAAACACCTTTCCTGGTAAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
 Db 361 TTGGCCCAACAACTCAAGAAAGAAATGCTGCTCCTTGTGATGTTGCGGCGGATATT 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 421 TATGCTCCAGCTGCCATTGACAGCTTAAGACCTTGGGACACAGATTGATGCTGCTGTC 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTGCGACTTGGAAACAGACAGTACAGCTGTTGAGATTGTCAGTCAAGGTTTGGCAAGCC 540  
 QY 181 LysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisLeuAspGlu 200  
 Db 541 CAATAATCATCAACGACTATGCTGTGATTGATCTCGGGTGTGTCAGATTGATGAG 600  
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 CTCCTCATGATGAGCTTGCTGATGTGAAGCATTGCTCAACCAATGAATCTGCTT 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
 Db 661 GTGCTGATGCTGATGTTGTCAGGAGCAGCAATGTTGCGGCTGAGTTTAATGCTCAG 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260  
 Db 721 TTGGAGTGACTGGGTGCATCTTACCAAGATTGATCGGCTACTCGTGGTGGTGCT 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 Db 781 CTGCTGTTGCTCACATTACTGAAAAACAATCAAGTTCACTGGTACAGGTGAAAAAGATT 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 ACGGACATTTAAACCTTCCACCCAGACCGCATGTCTACCGGTATCCTTGGTATGGGGAT 900  
 QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 Db 901 ATGCTCATTGTGATGGAAGCCTTCTCAGGAATACGATGAACAAAAGCCCTTGAAATG 960  
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340  
 Db 961 GCTGAGAAGATGCGGAAAAACACCTTTGATTTTAATGATTTATCATCATGATCAATAGATCAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 1021 GTGCAAAATATGGGCGGATGGAAGACTTGCTCAAGATGATCCAGGTATGCCCAACAT 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
 Db 1081 CCAGCCCTTCAAAACATGAGGTGGATGAACCGCAGATTGCTGCTAAACGTCGCAATGTG 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400

```

Db 1141 TCTTGATGACACCTGTAAGAGCGTGAACCCAGAGATTGTAAATCCAAAGCGTCGCGGT 1200
QY 401 ArgilleAlaLysGlySerGlyArgSerLeuGlnGlnValAlaAsnArgLeuMetLysGlnPhe 420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1201 CGTATGCTGCTGTTCTGGAATATCATTCGAGAGTCATTAATATCATCAGGACATTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1261 AACCCAGCTAAACAGCTCATGACGGGTGTATGCTGGG----- 1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1300 -----GATATCAATAAATGATGAGCAAAATGGGATTAAATCCAAATAACCTTCT 1350

RESULT 14
AAS55790
ID AAS55790 standard; DNA; 1395 BP.
XX
AC AAS55790;
XX
DT 13-FEB-2002 (first entry)
XX
DE Streptococcus pneumoniae DNA for cellular proliferation protein #361.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Streptococcus pneumoniae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-27931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU37931.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID NO 9427; 51lpp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic

```

```

CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1395 BP; 415 A; 272 C; 333 G; 375 T; 0 other;

Alignment Scores:
Pred. No.: 2,81e-109 Length: 1395
Score: 1306.00 Matches: 252
Percent Similarity: 74.73% Conservative: 91
Best Local Similarity: 54.90% Mismatches: 102
Query Match: 57.43% Indels: 14
DB: 23 Gaps: 2

US-09-943-108A-2 (1-455) x AAS55790 (1-1395)

QY 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGGCATTGAAAGTTTAAACAGAACGTTTGCAGAACGCTTTTAAATAATCTACGTAATAA 60
QY 21 GlyLysLeuThrGluAlaAspPheLysLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GGAATAATCTGATCTGATGTCGAGAGGCAACCAAGAAATTCGCTGGCCTGGTC 120
QY 41 GluAlaAspValAsnPheLysValValLysGluPheLysThrValSerGluArgAla 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GAGGCGGACGTTGCCTGCTGTTGTAAGGACTTATCAAGAAAGTTCGTGAGCGTGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysValGln 80
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GTCGGGCATGAGGTCATTGATACACTTAATCTCGCAACAGATTATTAATTCCTGAT 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GAGGAACCTGACAGCGCTTTAGGTTCTGATAGCGCAGAAATATCAAGTCACCTAAGATT 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 CCAACCATCATCATGATGTTGGTTTACAAAGGGGTGTAAACAACACCTTTGCTGTGATAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIle 140
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 TTGGCCCAACAACCTCAAGAAAGAAATGCTGCTCTTTGATGATTGGCGGGATATT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 TATGTCGACGTGCCATTGACCACTTAAAGACCTTGGGACACAGATGATGGCCTGTC 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TTTGCACCTTGGACAGAAAGTACCAGCTGTTGAGATTGTACGTCAAGGTTTGGAGCAAGCC 540
QY 181 LysGluGluHisLeuAspPheValIleAspThrAlaGlyArgLeuHisLysLysGlu 200
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 CAACATAATCATCAACGACTATGCTTGTATGATGCTCGGGTGGTGTGATGATGATGAG 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 CTCCTCATGATGAGCTTCGTGATGTAAGCATTTGGCTCAACCAATGNAATCTGGTT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAlaLagLusPheAspAspGln 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 GTCGTTGATGATGATGTTGTCAGGAAGCAGCAATGTTGCGGTGAGTTTAATGCTCAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 TTGGAAGTGAAGTGGGGTCATCTTACCAAGATTGATGGCGATACGCTGCTGCTGCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CAGTCTGTTGCTGACATTCTGGAACCAACCAAGTTCAAGTTCAAGTACAGGTGAAAGATT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
Db 601 CTCCTCATGATGAGTCTGCTGAGAAACATTCGCTCAACCAATGAAATCTGCTT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValValAlaGluSerPheAspGln 240
Db 661 GTCGTGATGCTGATGATGCTGAGAAACCAATTCGCGTGAGTTAAATGCTCAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 TTGGAGTGACTGGGGTCATCCTACCAAGATTGATGGCGATCTCGTGGTGGTCTGCT 780
QY 261 LeuSerLeArgSerValThrGlnLysProLeLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGTCTGTCTGCATCTACTGGAACCAATCAAGTTCACTGGTACAGGTCAAAAGATT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACGGCATTTGAACCTTCACCCAGACCGCATGCTAGCCGATCTCTGGTATGGGGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 ATGCTCACTTGATGAGAAGCTTCTCAGGAATACGATGACAAAGAACCTTCGAATG 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAGATCGCGAAACACCTTGATTTTAATGATTTCATCGATCAATTAGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGCAAAATATGGGGCGATGGGAAGACTTGCTCAAGATGATTCACGATATGCCAAACAAT 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCCCTTCAAAATGAGGTGGATGACGCCAGATTGCTCGTAAACGTCCCATTTGTG 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACACCTGAAGAGCGTGAAACCCAGATTGTTAAATCCAAGCGGTGCGCGT 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGTGCTGTTCTGGAATACATTCGCGAAGTCAATAAATTCATCAAGGACTTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys 440
Db 1261 AACCAAGCTAACAAGCTCATCGCGGTGTTATGCTGGG----- 1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATGATGAAGCAAAATGGGGATTATCCAAATAACCTTCT 1350
```

Search completed: February 25, 2003, 01:52:44  
Job time : 421 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 01:43:57 ; Search time 81 Seconds

(without alignments)  
1722.690 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274

Sequence: 1 MAFEGLSERLQATMCKMRGK.....GKKGRNQMNLMKGNLFF 455

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09943108/runat\_21022003\_151631\_1864/app-query.fasta\_1.647  
-DB=Issued\_Patents\_NA -OFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09943108.@cgnl\_1.61@runat\_21022003\_151631\_1864 -NCPU=6 -ICPU=3  
-NO\_XLIPX -NO\_MMALP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCUTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2274	100.0	1368	4	US-09-035-382-1
2	1534	67.5	1014	4	US-09-134-001C-1958
3	1306	57.4	7577	4	US-08-961-527-46
4	1305	57.4	792	4	US-09-035-382-3
5	1297	57.0	1569	2	US-08-923-772-1
6	1297	57.0	1569	4	US-09-385-287-1
7	1003.5	44.1	1098	4	US-09-221-017B-928
8	853	37.5	4804	4	US-09-066-047-6
9	592.5	26.1	2166	2	US-08-317-401B-3
10	573.5	25.2	2877	2	US-08-317-401B-1
11	432	19.0	4370	4	US-08-981-527A-20
12	430	18.9	1015	4	US-08-981-527A-7

13	426.5	18.8	1248	4	US-09-134-001C-1980
14	425	18.7	836	4	US-08-858-207A-180
15	421	18.5	1251	3	US-09-007-476-1
16	397.5	17.5	1278	2	US-08-986-963-1
17	397.5	17.5	1278	2	US-08-986-963-3
18	397.5	17.5	1293	3	US-09-007-484-1
19	397.5	17.5	1293	4	US-09-309-682-1
20	397.5	17.5	2764	2	US-08-986-963-4
21	396.5	17.4	13121	4	US-08-961-527-126
22	390	17.2	831	3	US-09-007-484-3
23	390	17.2	831	4	US-09-309-682-3
24	351	15.4	390	4	US-08-651-155B-4
25	344.5	15.1	4403765	4	US-09-103-840A-2
26	344.5	15.1	4411529	4	US-09-103-840A-1
27	329	14.5	1046	4	US-09-221-017B-496
28	267.5	11.8	29793	4	US-09-302-812-38
29	267.5	11.8	29793	4	US-09-511-477-38
30	267.5	11.8	29793	4	US-09-511-507-38
31	258	11.3	802	4	US-08-998-416-483
32	137.5	6.0	2085	2	US-08-668-128B-7
33	137.5	6.0	2085	2	US-08-905-445-7
34	136	6.0	1821	4	US-08-477-831C-1
35	136	6.0	1885	4	US-08-477-831C-9
36	136	6.0	1961	4	US-08-477-831C-8
37	136	6.0	2968	4	US-08-477-831C-13
38	136	6.0	3044	4	US-08-477-831C-12
39	135.5	6.0	1896	4	US-08-477-831C-10
40	132	5.8	3033	4	US-09-134-001C-2341
41	131	5.8	30549	4	US-09-134-001C-322
42	130.5	5.7	5361	4	US-08-973-462-2
43	130.5	5.7	6152	4	US-08-973-462-1
44	130	5.7	11091	4	US-09-134-001C-2243
45	128	5.6	3837	4	US-09-724-517-1

#### ALIGNMENTS

##### RESULT 1

US-09-035-382-1  
; Sequence 1, Application US/09035382  
; Patent No. 6284515  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael T.  
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES  
; FILE OF INVENTION: AND POLYNUCLEOTIDES  
; FILE REFERENCE: GM50035  
; CURRENT APPLICATION NUMBER: US/09/035,382  
; CURRENT FILING DATE: 1998-03-05  
; EARLIER APPLICATION NUMBER: 60/057,890  
; EARLIER FILING DATE: 1997-09-03  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1368  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-035-382-1

Alignment Scores:	2.03e-249	Length:	1368
Pred. No.:	2274.00	Matches:	455
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0
US-09-943-108A-2 (1-455) x US-09-035-382-1 (1-1368)			
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlyLysMetArgGlyLys 20			
Db 1 AFGGCAITTTGAAGGCTATCAAGACGCTGCAGCGACGACGATGCAAAAATGCGTGGTAAG 60			
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40			

```
Db 61 GCTAAACTTACTGAAGTCGATATAAGATAATGATGCGTCAAGTAAAGATTACGCTTATT 120
QY 41 GUAlaAspValAsnPhelYsValYsGluPhelLeuYsThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAACATTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGAACGGCCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleYsIleValGln 80
Db 181 TTAGGTTCCGATGTAATCAATCATACACAGGGCAACAAGTATTAAATAGTTCAA 240
QY 81 AspGluLeuThrYsLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTACGAAGTTGATGGTGGAGAAAAATACATCGATTAAATGTCAAATAACCA 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyYsThrThrAlaGlyLys 120
Db 301 CCTACTGTGTATGATGGTGTGTTACAAGGTGCTGGTAAAAACAACACTCGAGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTAGTGGTAAAAATACACAAAAACCTATGTAGTTCACACAGATATT 420
QY 141 TyrArgProAlaLalIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCATATAATCAATACAAACAGTAGGGAACAAATTCATATTCCTGTA 480
QY 161 TyrSerGluGlyAspGlnValYsProGlnGlnIleValThrAsnAlaLeuYsHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAGCCACACAAATTTGTAATAATCATTAATAAACATGCT 540
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACAATTAGACTTTGTATCATCTGATACACAGCGTCGATACACATCGAAGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValYsGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTGATGAACGAATTAAGAAGATAAAGAAATTCGTAACCAACCAAGAAATATTGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspIle 240
Db 661 GTGTGCGATTCAAGTACGGGTCAAGATGCTGTCAATGTGTCAATGCTTTTACCATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTACTTAACTAAATTAGATTGTTGATACACGTCGTCGTCACCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTATCTATTCTGTTGCGTGACAAAAACCAATTAATTTGTTGTTATGAGTGAAAAAGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetClyAsp 300
Db 841 GATGTTTAGACGATTTCCATCCCTGAAGTATGCGATCAGCTATTTTAGGTATGGGTGAT 900
QY 301 ValLeuSerLeuIleGlyLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTTTAAAGTTTAAATGAAAAAGCGCAACAAGATGTGGATCAAGAAAAACCAAAAGATT 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAAGATGCGTGTAGTCATCGTTTACTTTAGTATGATTTTAAAGCAACTTGTATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATCTAGACCCACTGGATGATATATGAAAAATGATTCAGGTATGAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGGCTTAGATAGCTTATATAGTGAAGCAAAATTCATATATTAAACGGATTATC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
|||||
```

```
Db 1141 CASTCAATGACCGCGCTGAAAGAAACAATCCAGACACATTAATGTATCATCGTAAAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CSTATTTGCTAAAGGCTTGTGTCGTTTCATTACAGAAGTCAATCGTTTGATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLysLysLysLys 440
Db 1261 AACGATATGAAGAAATGATGAACAATCTACTGGTGGCGGTAAAGTAAAGTAAGTAAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
|||||
Db 1321 CGCAATCAATGCAAAATATGTTAAAAAGTATGAATTTACCGTTT 1365
RESULT 2
US-09-134-001C-1958
; Sequence 1958, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1958
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1958
Alignment Scores:
Pred. No.: 2,35e-165 Length: 1014
Score: 1534.00 Matches: 307
Percent Similarity: 97.29% Conservative: 16
Best Local Similarity: 92.47% Mismatches: 9
Query Match: 67.46% Indels: 0
DB: 4 Gaps: 0
US-09-943-108A-2 (1-455) x US-09-134-001C-1958 (1-1014)
QY 1 MetalAPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 16 ATGGCATTTGAAGGATTTATCCGATCGCTTACAAGCCACGATGCAAAAAATGCGTGTAAA 75
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 76 GGAAGATGACGAGCAGATATTAACGATGATGCGTGAAGTGAATAGCGTTATTG 135
QY 41 GluAlaAspValAsnPhelYsValYsGluPhelLeuYsThrValSerGluArgAla 60
Db 136 GAAGCCGATGTTAACTCAAGGTGTGTAAGGAATTTGTTAAGAAATGTTTCAGAACGCG 195
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 196 CTAGGTTCTGATGTCATGCACTTAAACCTTTAAACCTGGCCAAACAGGTTTAAAAATCGTAA 255
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 256 GAAGAACTTACTAGTTTAAAGGTGGAGAAATATCTTCCATTAAGATGCAACCAACCA 315
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 316 CCAACTGTTGTCATGATGCTCGCTTACAAGGTGCAAGTAAACAGCAGCACACAGGTAAG 375
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 376 TTGGCATTTAATGCGTAAAAAATATAACAAAAAACCTTTACTTTGTCGAGGATATT 435
```

```
141 TyrArgProAlaAlaLeuGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
142 :
143 :
144 :
145 :
146 :
147 :
148 :
149 :
150 :
151 :
152 :
153 :
154 :
155 :
156 :
157 :
158 :
159 :
160 :
161 :
162 :
163 :
164 :
165 :
166 :
167 :
168 :
169 :
170 :
171 :
172 :
173 :
174 :
175 :
176 :
177 :
178 :
179 :
180 :
181 :
182 :
183 :
184 :
185 :
186 :
187 :
188 :
189 :
190 :
191 :
192 :
193 :
194 :
195 :
196 :
197 :
198 :
199 :
200 :
201 :
202 :
203 :
204 :
205 :
206 :
207 :
208 :
209 :
210 :
211 :
212 :
213 :
214 :
215 :
216 :
217 :
218 :
219 :
220 :
221 :
222 :
223 :
224 :
225 :
226 :
227 :
228 :
229 :
230 :
231 :
232 :
233 :
234 :
235 :
236 :
237 :
238 :
239 :
240 :
241 :
242 :
243 :
244 :
245 :
246 :
247 :
248 :
249 :
250 :
251 :
252 :
253 :
254 :
255 :
256 :
257 :
258 :
259 :
260 :
261 :
262 :
263 :
264 :
265 :
266 :
267 :
268 :
269 :
270 :
271 :
272 :
273 :
274 :
275 :
276 :
277 :
278 :
279 :
280 :
281 :
282 :
283 :
284 :
285 :
286 :
287 :
288 :
289 :
290 :
291 :
292 :
293 :
294 :
295 :
296 :
297 :
298 :
299 :
300 :
301 :
302 :
303 :
304 :
305 :
306 :
307 :
308 :
309 :
310 :
311 :
312 :
313 :
314 :
315 :
316 :
317 :
318 :
319 :
320 :
321 :
322 :
323 :
324 :
325 :
326 :
327 :
328 :
329 :
330 :
331 :
332 :
333 :
334 :
335 :
336 :
337 :
338 :
339 :
340 :
341 :
342 :
343 :
344 :
345 :
346 :
347 :
348 :
349 :
350 :
351 :
352 :
353 :
354 :
355 :
356 :
357 :
358 :
359 :
360 :
361 :
362 :
363 :
364 :
365 :
366 :
367 :
368 :
369 :
370 :
371 :
372 :
373 :
374 :
375 :
376 :
377 :
378 :
379 :
380 :
381 :
382 :
383 :
384 :
385 :
386 :
387 :
388 :
389 :
390 :
391 :
392 :
393 :
394 :
395 :
396 :
397 :
398 :
399 :
400 :
401 :
402 :
403 :
404 :
405 :
406 :
407 :
408 :
409 :
410 :
411 :
412 :
413 :
414 :
415 :
416 :
417 :
418 :
419 :
420 :
421 :
422 :
423 :
424 :
425 :
426 :
427 :
428 :
429 :
430 :
431 :
432 :
433 :
434 :
435 :
436 :
437 :
438 :
439 :
440 :
441 :
442 :
443 :
444 :
445 :
446 :
447 :
448 :
449 :
450 :
451 :
452 :
453 :
454 :
455 :
456 :
457 :
458 :
459 :
460 :
461 :
462 :
463 :
464 :
465 :
466 :
467 :
468 :
469 :
470 :
471 :
472 :
473 :
474 :
475 :
476 :
477 :
478 :
479 :
480 :
481 :
482 :
483 :
484 :
485 :
486 :
487 :
488 :
489 :
490 :
491 :
492 :
493 :
494 :
495 :
496 :
497 :
498 :
499 :
500 :
501 :
502 :
503 :
504 :
505 :
506 :
507 :
508 :
509 :
510 :
511 :
512 :
513 :
514 :
515 :
516 :
517 :
518 :
519 :
520 :
521 :
522 :
523 :
524 :
525 :
526 :
527 :
528 :
529 :
530 :
531 :
532 :
533 :
534 :
535 :
536 :
537 :
538 :
539 :
540 :
541 :
542 :
543 :
544 :
545 :
546 :
547 :
548 :
549 :
550 :
551 :
552 :
553 :
554 :
555 :
556 :
557 :
558 :
559 :
560 :
561 :
562 :
563 :
564 :
565 :
566 :
567 :
568 :
569 :
570 :
571 :
572 :
573 :
574 :
575 :
576 :
577 :
578 :
579 :
580 :
581 :
582 :
583 :
584 :
585 :
586 :
587 :
588 :
589 :
590 :
591 :
592 :
593 :
594 :
595 :
596 :
597 :
598 :
599 :
600 :
601 :
602 :
603 :
604 :
605 :
606 :
607 :
608 :
609 :
610 :
611 :
612 :
613 :
614 :
615 :
616 :
617 :
618 :
619 :
620 :
621 :
622 :
623 :
624 :
625 :
626 :
627 :
628 :
629 :
630 :
631 :
632 :
633 :
634 :
635 :
636 :
637 :
638 :
639 :
640 :
641 :
642 :
643 :
644 :
645 :
646 :
647 :
648 :
649 :
650 :
651 :
652 :
653 :
654 :
655 :
656 :
657 :
658 :
659 :
660 :
661 :
662 :
663 :
664 :
665 :
666 :
667 :
668 :
669 :
670 :
671 :
672 :
673 :
674 :
675 :
676 :
677 :
678 :
679 :
680 :
681 :
682 :
683 :
684 :
685 :
686 :
687 :
688 :
689 :
690 :
691 :
692 :
693 :
694 :
695 :
696 :
697 :
698 :
699 :
700 :
701 :
702 :
703 :
704 :
705 :
706 :
707 :
708 :
709 :
710 :
711 :
712 :
713 :
714 :
715 :
716 :
717 :
718 :
719 :
720 :
721 :
722 :
723 :
724 :
725 :
726 :
727 :
728 :
729 :
730 :
731 :
732 :
733 :
734 :
735 :
736 :
737 :
738 :
739 :
740 :
741 :
742 :
743 :
744 :
745 :
746 :
747 :
748 :
749 :
750 :
751 :
752 :
753 :
754 :
755 :
756 :
757 :
758 :
759 :
760 :
761 :
762 :
763 :
764 :
765 :
766 :
767 :
768 :
769 :
770 :
771 :
772 :
773 :
774 :
775 :
776 :
777 :
778 :
779 :
780 :
781 :
782 :
783 :
784 :
785 :
786 :
787 :
788 :
789 :
790 :
791 :
792 :
793 :
794 :
795 :
796 :
797 :
798 :
799 :
800 :
801 :
802 :
803 :
804 :
805 :
806 :
807 :
808 :
809 :
810 :
811 :
812 :
813 :
814 :
815 :
816 :
817 :
818 :
819 :
820 :
821 :
822 :
823 :
824 :
825 :
826 :
827 :
828 :
829 :
830 :
831 :
832 :
833 :
834 :
835 :
836 :
837 :
838 :
839 :
840 :
841 :
842 :
843 :
844 :
845 :
846 :
847 :
848 :
849 :
850 :
851 :
852 :
853 :
854 :
855 :
856 :
857 :
858 :
859 :
860 :
861 :
862 :
863 :
864 :
865 :
866 :
867 :
868 :
869 :
870 :
871 :
872 :
873 :
874 :
875 :
876 :
877 :
878 :
879 :
880 :
881 :
882 :
883 :
884 :
885 :
886 :
887 :
888 :
889 :
890 :
891 :
892 :
893 :
894 :
895 :
896 :
897 :
898 :
899 :
900 :
901 :
902 :
903 :
904 :
905 :
906 :
907 :
908 :
909 :
910 :
911 :
912 :
913 :
914 :
915 :
916 :
917 :
918 :
919 :
920 :
921 :
922 :
923 :
924 :
925 :
926 :
927 :
928 :
929 :
930 :
931 :
932 :
933 :
934 :
935 :
936 :
937 :
938 :
939 :
940 :
941 :
942 :
943 :
944 :
945 :
946 :
947 :
948 :
949 :
950 :
951 :
952 :
953 :
954 :
955 :
956 :
957 :
958 :
959 :
960 :
961 :
962 :
963 :
964 :
965 :
966 :
967 :
968 :
969 :
970 :
971 :
972 :
973 :
974 :
975 :
976 :
977 :
978 :
979 :
980 :
981 :
982 :
983 :
984 :
985 :
986 :
987 :
988 :
989 :
990 :
991 :
992 :
993 :
994 :
995 :
996 :
997 :
998 :
999 :
1000 :
```

```
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 7577 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-46

Alignment Scores:
Pred. No.: 4,91e-138 Length: 7577
Score: 1306.00 Matches: 252
Percent Similarity: 74.73% Conservative: 91
Best Local Similarity: 54.90% Mismatches: 102
Query Match: 57.43% Indels: 14
DB: 4 Gaps: 2

US-09-943-108a-2 (1-455) x US-08-961-527-46 (1-7577)
QY 1 MetaAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 3033 ATGCAATTCGAAAGTTTAAACAGACGTTTCGACAGCGCTTTTAAAAATCTACGTAATAAAA 2974
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 2973 GGAAAAATCTGTAATCTGATGTCCTCAAGAGGCAACCAAGAAATTCGCTGGCGCTGCTC 2914
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 2913 GAGCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2854
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 2853 GTCGGCGATGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2794
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 2793 GAGGATGACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2734
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 2733 CCAACCATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2674
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 2673 TTGCCCAACAACCTCAAGAAAGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2614
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 2613 TATGCTCCAGCTGCCATTGACCACTTAAGACCTTGGGACACACAGATTGATGCTGCTGCTGCT 2554
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 2553 TTTCGACTTGGACAGAAAGTACCAGCTGTTGAGATGTTACGTTCAAGGTTTGGAGCAAGCC 2494
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 2493 CAAACTAATCATACGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2434
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 2433 CTCCTCATGATGAGCTTCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2374
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 2373 GTCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2314
QY 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 2313 TTGGAAGTGACTGGGGTCACTCTTACCAGAGATGATGGCGATCTCTGCTGCTGCTGCTGCTGCT 2254
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
```

Db	2253	CTGCTGTTGGTCACATTACTGAAACCAATCAAGTCACCTGCPACAGGTGAAAGATT	2194
QY	281	AspGlyLeuGluPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
Db	2193	ACGGACATTGAACCTTCACCCACGACGCGATGCTAGCGCTATCCTTGGTATGGGGAT	2134
QY	301	ValLeuSerLeuIleGluLysAlaGlnInLaspValAspGlnGluLysAlaLysAspLeu	320
Db	2133	ATGCTCAGCTTTGATTGAGAAGCTTCTCAGGAATPACATGACACAAAAGCCCTTGAATG	2074
QY	321	GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGluLeuAspGln	340
Db	2073	GCTGAGAGATGCGGCAAAACACCTTGATTTATGATGTTCTGCGATCAATTGATCAG	2014
QY	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
Db	2013	GTGCAAAATATGGCGCCATGGAGAACTTGCTCAAGATGATTCACAGGTATGGCCACAAT	1954
QY	361	LysGlyLeuAspLysLeuAsnMetSerGlnLysGlnIleAspHisIleLysAlaIleIle	380
Db	1953	CAAGCCCTCAAAACATGAAGGTGGATGAAGCCGACATGCTGCGPAAACGTGCCATTGTG	1894
QY	381	GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys	400
Db	1893	TCTTCGATGACACCTTGAGAGCGTGAACCCAGATTGTGTTAAATCCACGCGTCGCGT	1834
QY	401	ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe	420
Db	1833	GCTATTGCTGCTGGTCTTGGAATACATTCGTCGAAGTGCATAAATTCATCAAGACCTT	1774
QY	421	AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys	440
Db	1773	AACAGGCTAAACAGCTCATGACAGGTGTTATGTCGGG-----	1735
QY	441	ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro	454
Db	1734	-----CATATGAATAAATGATGAAGCAAAATGGGGATTAATCCAAAATAACCTTCT	1684

Qy	21	GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
Db	61	GGTAAACCTTACTGAAGCTGATATACGATAAAGTAAGTGCCTGAAGTAAGATTAGCGTTATTT	120
Qy	41	GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla	60
Db	121	GAGGCTGACGTAAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCGAAGCGCA	180
Qy	61	LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln	80
Db	181	TTAGGTCCTCGATCATATGCAATCATTAACACAGCGGCAACAGTATTAAATAGTCAA	240
Qy	81	AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLleAsnMetSerAsnLysPro	100
Db	241	GATGAATTAACGAGTTGATGGTGGAGAAAATACATCGATTAAATGTCAAAATAAACCA	300
Qy	101	ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys	120
Db	301	CCTACTGTGTGTATGATGTTGGTTTACAGTGTCTGTGTAACACAACTCTCGAGTAA	360
Qy	121	LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIle	140
Db	361	TTAGCATATTGTCGTAATAAATAACACAAAACCTATGTTAGTTCACACAGATATT	420
Qy	141	TyrArgProIaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
Db	421	TATCTCTCAGCAGCGATAAATCAATTTACAAACAGTAGGGAACAAATTCATTCCTGTA	480
Qy	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysIleAla	180
Db	481	TPAGTGAAGAGATCAAGTAAAGCCACACAAATTTGAATGCTTTAAACACATGCT	540
Qy	181	LysGlnGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu	200
Db	541	AAAGAAGACAACTTACAGCTTGTATCATGTATGATACAGCGTCGATACACATCATGAA	600
Qy	201	AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu	220
Db	601	GCATGTGATGACGAATTAAGAAGTAGTAAAGAAATGTCTAAACCAACAGAAATATGTTA	660
Qy	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln	240
Db	661	GTGTGCGATTCAATGACGGGTCAGATGCTGTGTCATGTTCAGAACTCTTTGACGATCAA	720
Qy	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla	260
Db	721	CTTGATGTCACAGGTGTACCTTAACTTAATTAGATGGTGATACCCGTGGTGGACGCT	780
Qy	261	LeuSerIleArg	264
Db	781	TTATCTATTGCT	792

```

RESULT 5
US-08-923-772-1
; Sequence 1, Application US/08923772
; Patent No. 5972651
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: NOVEL f/fh
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; Zip: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

```



; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/923,772  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Dickinson, Todd Q  
 ; REGISTRATION NUMBER: 28,354  
 ; REFERENCE/DOCKET NUMBER: GM10080  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-994-2252  
 ; TELEFAX: 215-994-2222  
 ; TELEX:

; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1569 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-923-772-1

## Alignment Scores:

Pred. No.: 4,49e-138 Length: 1569  
 Score: 1297.00 Matches: 251  
 Percent Similarity: 74.51% Conservative: 91  
 Best local Similarity: 54.68% Mismatches: 103  
 Query Match: 57.04% Indels: 14  
 DB: 2 Gaps: 2

US-09-943-108a-2 (1-455) x US-08-923-772-1 (1-1569)

QY 1 MetAlaPheGluGlyLeuSerGluGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGGCAATTTGAAAGTTTAAACAGACGTTTCAGAACGCTTTTAAATAATCTACGTAAATAA 60  
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgLysValArgLeuAlaLeuPhe 40  
 Db 61 GAAATAATCTGATCTGATCCAGAGGACACCAAGAAAGTTCGGTGGCTGCTC 120  
 QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 121 GAGCCGACGTTGCCCTGCTGTTAAAGGACTTTATCAAGAAAGTTCGGTGGCTGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 Db 181 CTCGGGATGAGTCAATTGATACACTTAATCCTCGCAACAGATTATTAATAATCGTTGAT 240  
 QY 81 AspGlnLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 241 GAGGACCTGACAGCCGTTTAGTCTGATACGCGCAGAAATATATCAAGTCAACCTAAGATT 300  
 QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCAACATCATCATGATGTTGTTTACAAAGGGCTGGTAAACACCTTTGCTGCTGATAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspIle 140  
 Db 361 TTGGCCCAACAACTCAAGAAAGAAATGCTGCTCTTTGATGATTGCGCGCGATATT 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 421 TATCTCCAGCTGCCATTGACCATTAAGACCTTGGGACACAGATTGATGCTGCTGTC 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTTCGACTTGGACAGAAGTACCAGCTGTGAGATTGTAGCTCAAGGTTTGGAGCAAGCC 540  
 QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuLysIleAspGln 200  
 Db 541 CAAACTAATCATACGACTATGCTTGATGATTACTCGGGCTGTTGACGATTGATGAG 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 CTCTCATGAATGAGCTTCGGTGATGAAAGTATGGCTCAACCAATGAATCTGCTT 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
 Db 661 GTCGTTGATGCTATGATTGGTCAGGAAGCAGCAATGTTGGCGGTGAATTAATGCTCAG 720  
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 721 TTGGAAGTACTGGGGTCACTCTTACCAAGATGATGTTGATCTGCTGCTGCTGCT 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280  
 Db 781 CTGCTGTTGCTCATCATCTGGGAACCAATCAAGTTCATGCTACAGTGAAGAAAT 840  
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 ACAGATATCGAAACCTTCCACCAGACCGTATGCTAGCCGTATCTTGGCATGGGGAT 900  
 QY 301 ValLeuSerIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
 Db 901 ATGCTCACTTGTATGAGAAAGCTTCAGGAATACGATGACAAAGAGCCCTTGAAGT 960  
 QY 321 GluLysLysMetArgLysSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
 Db 961 GCTGAGAAGATGCGGAAACACCTTTGATTTAATGATTTTCATCGATCAATTAGATCAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 1021 GTGCAGAAATATGGGCGCATGGAAGACTTCTCAAGATGATTCAGGTTATGCCCAACAAT 1080  
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
 Db 1081 CCAGACTCTCAAAACATGAAGTGGTGAACGCCAGATTGCTGTAACGTCGCATTTG 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 TCTTCGATGACATCTCAAGACGTCGAAACCCAGATTGTTAAATCCCAAGCGTCGCGCT 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 CGTATTGCTGCTGTTGGAATACATCTGTCGAAGTCAATAAATTCATCAAGGACTTT 1260  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLysGlyLys 440  
 Db 1261 AACGAGCTTAACAGCTCATGCGAGGCTGTTATGCTCTGGG----- 1299  
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454  
 Db 1300 -----GATATGAATAAATGATGAAGCAAAATGGGATTAAATCCAAATACCTTCCT 1350

## RESULT 6

US-09-385-287-1  
 ; Sequence 1, Application US/09385287  
 ; Patent No. 6350857  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Black, Michael T.  
 ; TITLE OF INVENTION: NOVEL fth  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dechert Price & Rhoads  
 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: US  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/385,287  
 FILING DATE: 30-Aug-1999  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/923,772  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Dickinson, Todd Q  
 REGISTRATION NUMBER: 28,354  
 REFERENCE/DOCKET NUMBER: GM10080  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-994-2252  
 TELEFAX: 215-994-2222  
 TELEX: <Unknown>  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1569 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-385-287-1

Alignment Scores:  
 Pred. No.: 4,49e-138 Length: 1569  
 Score: 1297.00 Matches: 251  
 Percent Similarity: 74.51% Conservative: 91  
 Best Local Similarity: 54.68% Mismatches: 103  
 Query Match: 57.04% Indels: 14  
 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-385-287-1 (1-1569)

QY 1 MetalApehGluLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 1 ATGCCATTTGAAAGTTTACAGAACGTTTCGAGAACCTCTTAAACATCAGTAAAGAA 60  
 QY 21 GlyLysLeuThrGluAlaAspLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 61 GGAATAATCTCTGAATCTGATCTCCAGAGCAACCAAGAAATTCGCTGGCTGCTC 120  
 QY 41 GluAlaAspValAsnPhelYsValLysGluPheIleLysThrValSerGluArgAla 60  
 Db 121 GAGGCCAGCATGCTGCTTGTCTTGAAGACTTTATCAAGAAAGTTTCGTGACGTGCA 180  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
 Db 181 GTCGGCATGAGTCTATTGATACCTTAATCTCGCAACAGATTATTAATTCGTTGAT 240  
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
 Db 241 GAGGAACACAGCCGTTTATAGTCTGTACGCGCAGAAATTTATCAAGTCACCTAAGATT 300  
 QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 301 CCAACATCATCATGATGTTGGTTTACAGGGCTGTGTAACACACCTTTGCTGGTAAA 360  
 QY 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspIle 140  
 Db 361 TTGGCCAAACAACTCAGAAAGAGAAATGCTGCTCTTTGATGATTCGCGCGGAGATT 420  
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
 Db 421 TATCGTCAGTCGCAATGACAGCTTAAGACCTTGGCAACAGATTGATGTCCTGTC 480  
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
 Db 481 TTTCGACTTGAACAGAGATACAGCTGTGTGAGATTCTAGCTCAAGGTTTGAGCAAGCC 540  
 QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200  
 Db 541 CAAATATCATACAGACTATCTCTGATTGATCTACTCGGGTCTGTTTGCAGATTGATGAG 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
 Db 601 CTCCTCATGATGAGCTTCGTGATGGAAGATTGCTCAACCAATGAATCTTGCTT 660  
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
 Db 661 GTCGTTGATGCTATGATGTCAGGAAGCAGCAATGTCGCGTGAGTTAATGCTCAG 720  
 QY 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
 Db 721 TTGGAAGTGATGGGGTCATCCTTACCAAGATTGATGGTGATCTCGTGGTGGTGGT 780  
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
 Db 781 CTGCTGTTGCTCACATCACTGGAAACCAACATCAAGTTCACCTGACAGGTGAAAAATT 840  
 QY 281 AspGlyLeuGlnLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
 Db 841 ACAGATATCGAAACCTTCCACCCAGACCGTATGCTAGCCGTATCCTTGCCATGGGGAT 900  
 QY 301 ValLeuSerLeuIleGlnLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
 Db 901 ATGCTCACTTGTGATGAGAAAGCTTCTCAGGAATACGATGACAAAGACCTTGAAATG 960  
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340  
 Db 961 GCTGAGAAGATGCGGAAACACCTTTGATTTTAAATGATTCATGATCAATAGATCAG 1020  
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
 Db 1021 GTCAAAATATGGGCGGATGGAAGACTTGTCAAGATGATCCAGTATGCGCCACCAAT 1080  
 QY 361 LysGlyLeuAspLysLysMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
 Db 1081 CCAGCACTTCAACATGAGGATGAGTGAACCCAGATTGCTCGTAAACGTCCTATTGGT 1140  
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
 Db 1141 TCTTCGATGACATCTGAAGAAGCTGAAACCCAGATTGTTAAATCCAAAGCGTCGCGT 1200  
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
 Db 1201 CGATTGCTGCTGCTGCTGGAATACATTCGTCGAGTCAATAAATTCATCAAGACTTT 1260  
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLysLysLys 440  
 Db 1261 AACAGGCTAAACAGCTCATGAGGGGTGTTATGCTGGG----- 1299  
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454  
 Db 1300 -----GATATGAATAAATGATGAAGCAANTGGGATTATCCAAATAACCTTCT 1350

# RESULT 7

US-09-221-017B-928/c  
 ; Sequence 928, Application US/09221017B  
 ; Patent No. 6444799  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ROSS, Bruce C.  
 ; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 1120  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 755 PACE MILL ROAD  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304-1018  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: Windows  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0b  
 ; CURRENT APPLICATION DATA:

```

/ APPLICATION NUMBER: US/09/221,017B
/ FILING DATE: 23-DEC-1998
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: Pp1182
/ FILING DATE: 31-DEC-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: Pp1546
/ FILING DATE: 30-JAN-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: Pp2911
/ FILING DATE: 09-APR-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/AU98/01023
/ FILING DATE: 10-DEC-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Monroy, Gladys H
/ REGISTRATION NUMBER: 32,430
/ REFERENCE/DOCKET NUMBER: 27340-20021.00
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-813-5600
/ TELEFAX: 650-494-0792
/ TELEX: 706141
/ INFORMATION FOR SEQ ID NO: 928:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1098 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: circular
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHEetical: NO
/ ANTI-SENSE: UNKNOWN
/ ORIGINAL SOURCE:
/ ORGANISM: PORPHYROMONAS GINGIVALIS
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 1..1098
/ US-09-221-017B-928

```

Alignment Scores:	6.81e-105	Length:	1098
Pred. No.:	1003.50	Matches:	190
Score:	73.17%	Conservative:	80
Percent Similarity:	51.49%	Mismatches:	96
Best Local Similarity:	44.13%	Indels:	3
Query Match:	4	Gaps:	2
DB:			
US-09-943-108A-2 (1-455) x US-09-221-017B-928 (1-1098)			
QY	17	MetArgGLyLysGLyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArg	36
Db	1098	CTCAAAGTGGAGGAGAAATTACCGAGATCAATGTACCGGAACCTCAAAGATGTAAGT	1039
QY	37	LeuAlaLeuPheGluAlaAspValasnPhelysValValLysGluPheIleLysThrVal	56
Db	1038	CGTGCCCTTTGGATGCGGACGTGAACATATCTGCTCCACAGCATTTACGGACTGGTG	979
QY	57	SerGluArgAlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValle	76
Db	978	AAGGAAAGAGCCATCGGCCAAACGTGTGACCTCCGTCGTCGGCGGAACGTGATGTG	919
QY	77	LysIleValGlnAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleasnMet	96
Db	918	AAGATCGTTCATGACCAATTGGCTGCTGTGATGGGGGACATCTGTGGATATTAAATCTC	859
QY	97	SerAsnLysProPThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThr	116
Db	858	AAAGGCAGT---CCGGCAGTAACTTCATGTCGGGATTCGAAGGTCCGGGTAGACACC	802
QY	117	ThrAlaGlyLysLeuAlaLeuLeuMetArgLysLysLysTrasnLysLysProMetLeuVal	136
Db	801	TTTCCTCGGCAAACTGGCCAAATATGCTTCAAGTCCAAAGCAGGCAACACGACCTTGCTCGTT	742

QY	137	AlaAlaAspIleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyIysGlnIle	156
Db	741	GCCTGCAGGCTGATCGCCTCGGGCTATCCAGCAGTGATGCTTCGGCAGCAGCTC	682
QY	157	AspIleProValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAla	176
Db	681	GGGTACCCGCTTACAGCGAAGCCGACACAAAAGCCGGTCGAGATAGCCGACAGCC	622
QY	177	LeuLysHisAlaLysGlnGluHisLeuAspPheValIleLeuAspThrAlaGlyLeu	196
Db	621	ATACCCGAGCCCAAGCAACCAAGGGACGACGTTGTCAATCGACACCCGSGTCGTTG	562
QY	197	HisIleAspGlnAlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsn	216
Db	561	GCCATCGACGAAGAATGATCGCGAGATAGAGGCCATCAAGGCTCCCATCCGCGAAC	502
QY	217	GluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSer	236
Db	501	GAACTCCTTTTCGTGTGGATCGATGACCGTCTAGGACGCGGTGAATACGCCAAGGAG	442
QY	237	PheAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArg	256
Db	441	TTCAATACACGCTTGATTCGACGGTGTTCCTTACCAAGCTCCGATGGAGATACCGG	382
QY	257	GlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMet	276
Db	381	GGCGGTGCCGCCCTCTCGATCCGTCGGTCGTAAACAAGCCCATCAAGTTCGTCGTACG	322
QY	277	SerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeu	296
Db	321	GGGAGAGATGGAGCAATCGATGCTTTCACCCCGAGCGTATGGCCGACCGTATCCTT	262
QY	297	GlyMetGlyAspValIleuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLys	316
Db	261	GGCATGGGGACATCGATCCCTTGTGGAGCGTGCTCAGGAGCAATACGACGACGGAA	202
QY	317	AlaLysAspLeuGluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlu	336
Db	201	GCTCGCAACTGGAGGAGAAATAGCAAGAATAGTTCGACTTCATGACTTCCTGCCC	142
QY	337	GlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGly	356
Db	141	CAGATACATCAGATCAAAAGATGGGTACCTCAAGAGCTTGCAWCCATGATCCCGGA	82
QY	357	MetAsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIle	376
Db	81	GTGGGA-----AGGCCATCAAGATATCGACATAGAGCAGCATGCC TTC AAGGATC	28
QY	377	LysAlaIleIleGlnSerMetThrPro	385
Db	27	GAGGCTATCATATATTTCATGACGCC	1

RESULT 8  
US-09-066-047-6  
; Sequence 6, Application US/09066047A  
; Patent No. 6306394  
; GENERAL INFORMATION:  
; APPLICANT: MURPHY, Cheryl  
; STOREY, James  
; BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF  
; USE OF GRANULOCYTIC ERLICHTIA  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,047A
; FILING DATE: 24-Apr-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/044,869
; FILING DATE: 25-Apr-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Superko, Colleen
; REGISTRATION NUMBER: 39,850
; REFERENCE/DOCKET NUMBER: 106.941.156
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-066-047-6

```

```

Alignment Scores:
Pred. No.:          9,27e-87      Length:      4804
Score:              853.00        Matches:     174
Percent Similarity: 63.57%        Conservative: 100
Best Local Similarity: 40.37%     Mismatches:  147
Query Match:        37.51%        Indels:      10
DB:                  4            Gaps:           3

```

US-09-943-108A-2 (1-455) x US-09-066-047-6 (1-4804)

```

QY 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 3490 CTTATGTTTAACTCTTAAACCAAGGGGTTTCTTCGCGTCAAAAGGTTAAGTGAAG 3549
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 3550 CCGGAGATATCCACGAGGATTGTGATCTGTATAGAGATATATCTCAGGCATTGTG 3609
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 3610 GATCGGATGTTAATCTGTGTGTGTGACCAATTTATAGAACGTAATAAAGCAAGATC 3669
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 3670 GTAGGGCGGATGATGATTAAGGGGTCCTCCGGAGCAATGTCATTAAGCGTATGAA 3729
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 3730 GAGTCTTTGATTGAAGTTTATAGTAATAGAGAGAGCGCTCTGATCTTAAGGGAAGATT 3789
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrThrAlaGlyLys 120
Db 3790 COTCGAGTAATCATGATGTTGGGCTCAAGGTGTGTGAAGACTACTAACACAGTAA 3849
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 3850 GTTGCACTGAGGTTA---AGAGAGGATTCTAAACCCGTGTGATGCTCTTACAGCTA 3906
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 3907 TATCGTCTCGACGTCGAGAACAGCTGAAGGTTTGGCTGATGGAGTGTGTATAGACAGT 3966
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180

```

```

Db 3967 CTTCCCATCGTTGAGGAGCAAAACCACTTCATATTCCGAAGCGTCTATGAGGAAAGCG 4026
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaIleAspThrAlaIleAspGlu 200
Db 4027 AGGCTCAAGGGCAGCATGTGTCTTTGGATACAGCGGGCGCTTCGATCATCAATCAG 4086
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluValIleAlaLysProAsnGlnIleMetLeu 220
Db 4087 GACATGATAGATGAGCTGAAGTGTCTAAGAGAGAGGTATCACCAGCTGAATGTATTG 4146
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 4147 GTTGTAGACTCCCTTAATGGGCAAGATGCCCTACTATGTCGCAAGTTCATCAATGAGG 4206
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 4207 TTAGGCATTACTGGGACGATCTTACCAGGCGGATGGTGATCCTTAGGGGTGGTCTATC 4266
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlyLysLeu 280
Db 4267 TTGTCTATGAAGTTGGTTCCTGATGTCCTATAAGTTCATGTCTACGGGAGAGAGCCT 4326
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 4327 GAAGATTGGACGATTCTATCCTGATAGATAGCTCGTAGAATGTTAAATATGGGAGAT 4386
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 4387 GTCCCATCTCTTGTAGAAAAGCGGTGAGACGGTTGGCAAGGATACAAATTAATGAGCTA 4446
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
Db 4447 CAGCGGAAGGCCAAGAGGTTAAATTCGATTGGATGATCTTGTATTTCAGCTGAAAGCT 4506
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMet----- 357
Db 4507 TTGATATAAATGGGTGATGCTCTAATAATGAAGTTTATACCCGCTTCGGTAACGAT 4566
QY 358 -----AsnLysMetLysGlyLeuAspLysLeuAsnMetSerGlyLysGlnIleAspHis 375
Db 4567 ATAAACCAAGTTCGGGGGATA-----CCTGATGACAGCAAGTCTGACATG 4614
QY 376 IleLysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsn 395
Db 4615 TACATTCGGATTATTAATCAATGACGAAGCAGGAGGCGGCAATCCTGAGATACCTGAAT 4674
QY 396 ValSerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArg 415
Db 4675 GGTGCGAGGAAGCAAGGATAGCAGGAGGTCGGGAGTTAAGGTTGATGCTGTAATGCG 4734
QY 416 LeuMetLysGlnPheAsnAspMetLysLysMet 426
Db 4735 TTGCTAAGCAGTATATAATCAGATGAATTCGATA 4767

```

# RESULT 9

```

US-08-317-401E-3
; Sequence 3, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; APPLICANT: Yaver, Debbie Sue
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy

```

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,401E
; FILING DATE: 03-October-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4248,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2166 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; US-08-317-401E-3

```

```

Alignment Scores:
Pred. No.: 1,24e-57 Length: 2166
Score: 592.50 Matches: 156
Percent Similarity: 50.11% Conservative: 75
Best Local Similarity: 33.84% Mismatches: 211
Query Match: 26.06% Indels: 19
DB: Gaps: 9

```

US-09-943-108a-2 (1-455) x US-08-317-401E-3 (1-2166)

```

QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 86 ATGGCTTCAGGATCTCGGCGGCGAATCAACGCGCGGCAATGACCTGACTCGCTCC 145
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 146 AACAATTTGGACGAGAGGCGCTTTGATGACATGATTAAAGAGATCTCGCGCGCTGCTG 205
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 206 TCGCGCGAGCAACGTCGCGCTGTCAGTCCCTCCGCAAGTCGCAATCAACGCGCTG 265
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 266 AACTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 325
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 326 GATGAGCTGGTTCCCTGGTGTATCCCATGCGGAGCGCTTCCGCGCCCAAGAGGCGCG 385
QY 101 ProThrValMetMetValGlyLeuGlnGlyIleValGlyLysThrThrAlaGlyLys 120
Db 386 TCAACGTCATCAGTCTCGTGGTCTGCGAGGTCGCGGTAAACACCACTGTACCAAG 445
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLys-----LysPrometLeuValAla 137
Db 446 CTGGCC-----CGCCATCCAGATGCGCGGCTTCAAGACTGCGCTGCTGT 493
QY 138 AlaAspIleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAsp 157
Db 494 GCGGATACCTCCGCTGGTGTCTTCGACCACTGAAGCAGAATGCCAACCAAGCCCAAG 553
QY 158 IleProValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeu 177
Db 554 ATCCCTACTACGTTAGTTCAGCAACCGACCGCCCGCCATGTAGCAGCCGAGGTGTG 613
QY 178 LysHisAlaLysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHis 197
Db 614 GCCAAGTTCAAGAGGAGCGTTTCGAAATCATCATCATCGTCGATACCATGTCGTCACAAG 673

```

```

QY 198 IleAspGluAlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGlu 217
Db 674 CAGGAAGAAGAGGCTTTTCACCGAAATGACCCAGACCCGCGCTCACCCGCGCAG 733
QY 218 IleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPhe 237
Db 734 ACCATCTCTGCTCCGACGACGACCATCGGTGAGGTCCGCAAGCCCGCTCTGCTTC 793
QY 238 AspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGly 257
Db 794 AAGGCCACCGCAGACTCGGAGCCATCATCATCAACAGACGATGGTCACGCGCAGGT 853
QY 258 GlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSer 277
Db 854 GGTGTGCTATTTCGCGCGCGCACACACTCCCATTTATCTACCTCGGTACCGGT 913
QY 278 GluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGly 297
Db 914 GAGCACCTGATGGATCTGGGACGCTTTGAGCGGAGGCGCTTCATCCAGAAGCTCTCGGT 973
QY 298 MetGlyAspValLeuSerLeuIleGluLysAlaGlnGln---AspValAspGlnGluLys 316
Db 974 ATGGCGATATGGCTGGCTGGTAGACGCGTACAGCCGTACAGCCGTGACCAAGACTCTGCCTCC 1033
QY 317 AlaLysAspLeuGluLysLysMetArgGluSerSerPheThrIleAspAspPheLeuGlu 336
Db 1034 GCCAAGGAACCTACAAGCAATCTCCGAGGTATCTACAGCTGCTGCTGCTCCGCGAG 1093
QY 337 GlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGly 356
Db 1094 AACATCCACTCCATCATGAGTGGGAGCCCTCTCCAAGCTCTCCGCGCATGATCCCGGT 1153
QY 357 Met----AsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHis 375
Db 1154 CTATCCAACTGACCGCGGCTCGATGATGACGAGAGCGCTCCATGAAG---CTCCGTCGC 1210
QY 376 IleLysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsn 395
Db 1211 ATGATCTACATTCGACATGATGACGCGCGCGGCTCCGATGCG---GATGGCAGAT 1267
QY 396 Val-----SerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGln 411
Db 1268 TTCTCGCAACACCCAGCCGCTGCTGCTGCTGCGGAGCGGTACCACTCCGCTCGC 1327
QY 412 GluValAsnArgLeuMetLysGlnPheAsnAspMetLysLysMetLysGlnPheThr 431
Db 1328 GAAGTGGAGACCTGCTCTCCAGCACCGCATGATGGCGCGCATGGCCAGCGGTGTC--- 1384
QY 432 GlyGlyGlyLysGlyLysArgAsnGlnMetGlnAsnMetLeuLysGlyMet 451
Db 1385 ---GGTGGCAGAGAAGACAGATGACGCGC-----GCCAGACATGCTCAAGGCGGT 1435
QY 452 Asn 452
Db 1436 AAC 1438

```

# RESULT 10

```

US-08-317-401E-1
; Sequence 1, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; APPLICANT: Yaver, Debbie Sue
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

```



```
; SEQ ID NO 20
; LENGTH: 4370
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-08-981-527A-20

Alignment Scores:
Pred. No.: 6,96e-39 Length: 4370
Score: 432.00 Matches: 114
Percent Similarity: 53.25% Conservative: 66
Best Local Similarity: 33.73% Mismatches: 120
Query Match: 19.00% Indels: 38
DB: 4 Gaps: 10

US-09-943-108a-2 (1-455) x US-08-981-527A-20 (1-4370)
QY 4 GluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeu 23
Db 2351 GATTCGGTATCTGAAAGTTTAAAGTATGGCTTGAAAAACAAGAACTCTTCAAAC 2410
QY 24 ThrGluAlaAspPheLysLeuMetMetArgGluValArgLeuAlaLeuPheGlu----- 41
Db 2411 AAAGTGAATGATCTGTATCCCGTTACCGTAAAGTGGATGAGGATTTCTCGAAGAGCTT 2470
QY 42 -----AlaAspValasnPhelysValLysValLysGluPheille----- 53
Db 2471 GAAGAGTTCTTATCACCGGGATGTCGGTTTACACCGTTATGGAATTAATAGAGAG 2530
QY 54 -----LysThrValSerGluArgAlaLeuGlySerAspValMetGlnSerLeuThrPro 71
Db 2531 CTGAAAAAAGTCAACACGAG-----AATATTCAAGAT 2566
QY 72 GlyGlnGlnValIleLysIleValGlnAspGluLeuThrLysLeu-----MetGlyGly 89
Db 2567 CCAAGAGGATCAAGTCAAGTATCTGAGAACTGGTCGAGATTTATAACAGCGGAGAT 2626
QY 90 GluAsnThrSer-----IleAsnMetSerAsnLysProProThrValValMetValGly 108
Db 2627 GAGCAAAATTCAGACTGACATCCAGATCGGCGTTTAAACGTAATCTCTGCTAGGT 2686
QY 109 LeuGlnGlnAlaGlyLysThrThrAlaGlyLysLeuLysLeuLeuMetArgLysLys 128
Db 2687 GTAAACGGCGCGGAAAAACAACAGATCGGAAAGCTTGTCATAAATG---AAACAA 2743
QY 129 TyrAsnLysLysProMetLeuValAlaAlaAspPheLysArgProAlaAlaIleAsnGln 148
Db 2744 GAAGGAAATCTGTGTACTTCGCCCGGAGACATTTTACGCGGAGCCATTGAACAG 2803
QY 149 LeuGlnThrValGlyLysGlnIleAspPheProValTyrSerGluGlyAspGlnValLys 168
Db 2804 CTGGAAGTATGGGAGAGCGTACAGAGTGCCTGTCATTAAAGCAGCGCGCAGAGCGCAT 2863
QY 169 ProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGlnHisLeuAspPheVal 188
Db 2864 CCGGGGCTGTCACTACGATGCTGTCATCGCGAAGCAAGAAATCCGATGATTA 2923
QY 189 IleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGlnLeuLysGlu 208
Db 2924 ATTTGTGATACGGCAGCGGCTCTCAAAAACAAGTAAATCTCATGAAGAGCTTGAAAA 2983
QY 209 ValLysGluIleAlaLys-----ProAsnGluIleMetLeuValVal 222
Db 2984 GTAAACGTGTATCGAAGAGAGTTCCTGAGCTCGCATGAGGTGCTGCTGCCCTT 3043
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGlnLeuAsp 242
Db 3044 GATGCCACGACGCGCAAAATGCAATGCTCGAGCAAAAGAAATTTCTTAAGCAACAAT 3103
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyLysAlaAlaLeuSer 262
Db 3104 GTTACCGCATGTCTTTAAGCAAGCTTGACGTACGGCAAAAGCGGTATCTCTCGG 3163
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282

; SEQ ID NO 20
; LENGTH: 4370
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-08-981-527A-7

Alignment Scores:
Pred. No.: 1.22e-39 Length: 1015
Score: 430.00 Matches: 114
Percent Similarity: 53.29% Conservative: 64
Best Local Similarity: 34.13% Mismatches: 118
Query Match: 18.91% Indels: 38
DB: 4 Gaps: 10

US-09-943-108a-2 (1-455) x US-08-981-527A-7 (1-1015)
QY 4 GluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeu 23
Db 64 GATTCGGTATCTGAAAGTTTAAAGTATGGCTTGAAAAACAAGAACTCTTCAAAC 123
QY 24 ThrGluAlaAspPheLysLeuMetMetArgGluValArgLeuAlaLeuPheGlu----- 41
Db 124 AAAGTGAATGATCTGTATCCCGTTACCGTAAAGTGGATGAGGATTTCTCGAAGAGCTT 183
QY 42 -----AlaAspValasnPhelysValLysValLysGluPheille----- 53
Db 184 GAAGAGTTCCTTATCACCGCGGATGTCGGTTTACACCGCTTATGGAATTAATAGATGAG 243
QY 54 -----LysThrValSerGluArgAlaLeuGlySerAspValMetGlnSerLeuThrPro 71
Db 244 CTGAAAAAAGAGTCAACACGAG-----AATATTCAAGAT 279
QY 72 GlyGlnGlnValIleLysIleValGlnAspGluLeuThrLysLeu-----MetGlyGly 89
Db 280 CCAAGAGGATCAAGTCAAGTATTTCTGAGAACTGGTCGAGATTTATAACAGCGGAGAT 339
QY 90 GluAsnThrSer-----IleAsnMetSerAsnLysProProThrValValMetValGly 108
Db 340 GAGCAAAATTCAGACTGACATCCAGGATCGGCGTTTAAACGTAATCTCTCTGCTAGGT 399
```

QY 109 LeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArgLysLys 128  
 Db 400 GTAACGCGTCGGGAAACACACAGATCGGAAGCTTCTCATAAATG--AAACA 456  
 QY 129 TyrAsnLysLysProMetLeuValAlaAlaAspLeuYrgrProAlaAlaLeuGln 148  
 Db 457 GAAGAAATATCTGTACTGCGCGGAGACACTTTTAGCGGGAGCCATTGAACAG 516  
 QY 149 LeuGlnThrValGlyGlnLeuAspLeuProValTyrSerGluGlyAspGlnValLys 168  
 Db 517 CTGGAGATATGGGAGAGCGTACAGAGTGCCTGTCATTAAAGACGCGCAGACCAT 576  
 QY 169 ProGlnGlnLeuValThrAsnAlaLeuLysHisAlaLysLeuLufHisLeuAspPheVal 188  
 Db 577 CCGGGGCTGTCTATCATCTACGATGCTGTTCATGCTCGAAAGCAAGAAATCGCATATTA 636  
 QY 189 IleLeuAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLysGlu 208  
 Db 637 ATTGTGATACGGCAGCGGCTCTCCAAACAAAGTAAATCTCATGAAGAGCTTGAAAAA 696  
 QY 209 ValLysGluIleAlaLys-----ProAsnGluIleMetLeuValVal 222  
 Db 697 GTAAACGTGTATCGAAAGAGAGTTCCTGAGCTCCGCAATGAGGTGCTTGCCTT 756  
 QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242  
 Db 757 GATGCCACGACGCGCCAAATGATGGCTCAGCAAAAGAAATTTCTTAAGCAACAAT 816  
 QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAlaLeuSer 262  
 Db 817 GTTACCGCATGCTTTAAGCAAGCTTGACGTAGCGCAAAAGCGGTATCGTCTTGG 876  
 QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282  
 Db 877 ATTCCGACAGAGCTTCATCCCGGTTAACTAGTCGCTTTAGGAGAAAGTGTGATGAC 936  
 QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302  
 Db 937 CTTCAAGGAATTGATCCAGAA-----TCCTATGTGACGGACTC-----TTTCA 981  
 QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLys 316  
 Db 982 GATTAGTGGAAAGCCGAC-----GATTAGAGAAA 1014

RESULT 13  
 US-09-134-001C-1980  
 ; Sequence 1980, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 1980  
 ; LENGTH: 1248  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-1980

Alignment Scores:  
 Pred. No.: 4,2e-39 Length: 1248  
 Score: 426.50 Matches: 103  
 Percent Similarity: 54.57% Conservative: 76  
 Best Local Similarity: 31.40% Mismatches: 122  
 Query Match: 18.76% Indels: 27  
 DB: 4 Gaps: 8

US-09-943-108A-2 (1-455) x US-09-134-001C-1980 (1-1248)  
 QY 8 GluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeuThrGluAlaAsp 27  
 Db 298 CAAACTTCCAAAGAACAGTAAATAATTAAATGCTGATATAGAAAAGTGAACGAAGAT 357  
 QY 28 IleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAlaAspValAsnPheLys 47  
 Db 358 TCTCTCGAAGCTCTGGAGAAATG-----CTATTACTGGGACGTTGGTTTAAAT 408  
 QY 48 ValValLysGluPheIleLysThrValSerGluArgAlaLeuGlySerAspValMetGln 67  
 Db 409 ACCGTATGAAATTAACATGATGAGCTACGTACAGAAAGCACAAAGACGTATATACAAGAA 468  
 QY 68 SerLeuThrProGlyGlnGlnValIle--LysIleVal-----GlnAsp 81  
 Db 469 ACAGAGACTTAAGAGAGATTATAGTTGAGAAGATGTAGAAATCTATCATCAAGAGAC 528  
 QY 82 GluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProPro 101  
 Db 529 GATTATCTGAGCAATGAATATTGAAGATGGACGTTTAAAT----- 570  
 QY 102 ThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeu 121  
 Db 571 --GTCATACTGATGGTGTGTAATGGTGTGCGCAAAACAAACAATTTGGTAAATTA 627  
 QY 122 AlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyr 141  
 Db 628 GCTTATCTGTTATCAACAGAA--GGTAAAAAAGTAAGTTAGCCGCTGATACCTTC 684  
 QY 142 ArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyr 161  
 Db 685 AGAGCTGGAGCAATTCACAATTAACACTCTGGGAGAACGTTGTTGTTGAAGTTGTG 744  
 QY 162 SerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLys 181  
 Db 745 AGTCAAAACGAGAGTCTGACCTCGCAGCAGTAGTATGATGGGATTAATGCCGCAAA 804  
 QY 182 GluGluHisLeuAspPhePheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAla 201  
 Db 805 AATAAGCGGTAGATATTTAAATTTGTGATCTGCAGGACGCTTGCAAAATAATCTAAT 864  
 QY 202 LeuMetAsnGluLeuLysGluValLysGluIleAlaLys-----Pro 215  
 Db 865 TTAATGCAAGAGTTAGATATAAATGAAACGTTGATTAATCGTCAATACCTGATGCCGCC 924  
 QY 216 AsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlu 235  
 Db 925 CATGAAGCTTTATATGCTTGGATGCAACACTGCTCAAAATGCCTTTTCCACAGCGGT 984  
 QY 236 SerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThr 255  
 Db 985 TCATTTAAGGAAGTTACAATGTCTCAGGTATAGTTTAACTAATATTAGACGCTACTGCT 1044  
 QY 256 ArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGly 275  
 Db 1045 AAAGGGGCTATTGTATTAGCAATTCGAAATGAGTTACACATTCACATTAATATATGTTGT 1104  
 QY 276 MetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIle 295  
 Db 1105 TTAGCGCAAAAAGTGGATGACTTACAACCGTTTAAATCTCTGAA-----AGCTATGTA 1155  
 QY 296 LeuGlyMet--GlyAspValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGln 314  
 Db 1156 TATGGATTATTTGCTGATATCATAGACAAATGAGATATTCTCGAAGAAATCTCTAGA 1215  
 QY 315 GluLysAlaLysAspLeuGluLys 322  
 Db 1216 AATTCATCGTTGAATCTGAAGAA 1239

RESULT 14  
 US-08-858-207A-180



; Sequence\*180, Application US/08858207A  
 ; Patent No. 6348328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Black, Michael  
 ; APPLICANT: Hodgson, John  
 ; APPLICANT: Knowles, David  
 ; APPLICANT: Nicholas, Richard  
 ; APPLICANT: Stodola, Robert  
 ; TITLE OF INVENTION: No. 6348328el Compounds  
 ; NUMBER OF SEQUENCES: 552  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406-0939  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/858,207A  
 ; FILING DATE: 09-MAY-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/017670  
 ; FILING DATE: 14-MAY-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gimmi, Edward R  
 ; REGISTRATION NUMBER: 38,891  
 ; REFERENCE/DOCKET NUMBER: P50475  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-270-4478  
 ; TELEFAX: 610-270-5090  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 180:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 836 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-858-207A-180

Alignment Scores:  
 Pred. No.: 3,34e-39 Length: 836  
 Score: 425.00 Matches: 83  
 Percent Similarity: 76.35% Conservative: 30  
 Best Local Similarity: 56.08% Mismatches: 35  
 Query Match: 18.69% Indels: 0  
 DB: 4 Gaps: 0

US-09-943-108a-2 (1-455) x US-08-858-207A-180 (1-836)

QY 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
 Db 393 ATGGCATTTGAATTTACAGACAGCTTTCAGACAGCTTTAAATAATCTAGCTAATAAAA 452  
 QY 21 GlyLysLeuThrGluAlaAspLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40  
 Db 453 GGAAAAATCTCTGAATCTGATGTCCTCAAGAGGCAACCAAGAAATTCGCTGGCCTTGCTC 512  
 QY 41 GluAlaAspValAsnPheLysValLysGluPheLysThrValSerGluArgAla 60  
 Db 513 GAGCGGACGTTGCTTGTGTAAAGGACTTTATCAAGAAAGTTCGTGAGCGGCA 572  
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLeuValGln 80  
 Db 573 CTCGGCGATGAGGTCATGTATACACTTAATCTCGGCAACAGATTTTAAATTCGTTGAT 632  
 QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerLeuAsnMetSerAsnLysPro 100  
 Db 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerLeuAsnMetSerAsnLysPro 100

Db 633 GAGGAACGTACAGCCGTTTTAGGTTCTGTATACGACGCAAAATATCAAGTCACCTAGATT 692  
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
 Db 693 CCAACCATCATCATGATGTTGTTTACAGAGGGCTGGTAAACACACCTTTGCTGGTAAA 752  
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAsp 140  
 Db 753 TTGCCCAACAACTCAAGAAAGAAATGCTCGTCCCTTGTGATGATTCGCGCGATATT 812  
 QY 141 TyrArgProAlaAlaLeuAsnGln 148  
 Db 813 TATGTCACCTGCCATTGACCAG 836

RESULT 15  
 US-09-007-476-1  
 ; Sequence 1, Application US/09007476  
 ; Patent No. 6159949  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Black, Michael T.  
 ; TITLE OF INVENTION: No. 6159949el Ftsy  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dechert Price & Rhoads  
 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: US  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/007,476  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Falk, Stephen T  
 ; REGISTRATION NUMBER: 36,795  
 ; REFERENCE/DOCKET NUMBER: GM10079  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-994-2488  
 ; TELEFAX: 215-994-2222  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1251 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-09-007-476-1

Alignment Scores:  
 Pred. No.: 1.78e-38 Length: 1251  
 Score: 421.00 Matches: 101  
 Percent Similarity: 55.59% Conservative: 78  
 Best Local Similarity: 31.37% Mismatches: 121  
 Query Match: 18.51% Indels: 22  
 DB: 3 Gaps: 7

US-09-943-108a-2 (1-455) x US-09-007-476-1 (1-1251)

QY 8 GluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeuThrGluAlaAsp 27  
 Db 289 CAAATTTCCAGAACAAATTAATATTGATAGCGAGATATCGTAAAGTAGATGAGAC 348  
 QY 28 IleLysIleMetMetArgGluValArgLeuPheGluAlaAspValAsnPhelys 47  
 Db 28 IleLysIleMetMetArgGluValArgLeuPheGluAlaAspValAsnPhelys 47

```
Db 349 TTTTGTGAAGCTTTAGAGAAATG-----TTATCACTCGAGACGTCGGTTTTTAAT 399
QY 48 valvalLysGluPheIleLysThrValSerGluArgAlaLeuGlySerAspValMetGln 67
Db 400 ACAGTG-----ATCAGCTTAATGAAGAATTAAGTATGGAAGCACACACGACGT 447
QY 68 SerLeuThrProGlyGlnValIleLysIleValGlnAspGluLeuThrLysLeuMet 87
Db 448 AATATTCAGATACATGAGATTTCGCGTAGTCATGTGAAAAAATCCTAGAGATTAC 507
QY 88 GlyGlyGlu-----AsnThrSerIleAsnMetSerAsnLysProProThrValVal 104
Db 508 CATCAAGAAGATGATATTCAGAAGCTATGAACATAGAGAGTGTGTTAAATGTCATT 567
QY 105 MetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeu 124
Db 568 TTAATGGTGGTGTGAATGTGTGTGTAACACACAAATTCGAAAAATTAGCT---TAC 624
QY 125 MetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArgProAla 144
Db 625 CGATATAAATCGAAGGTAAAAAGTAATGTTAGCTGGCGCGATACTTTAGACGGGT 684
QY 145 AlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGly 164
Db 685 GCTATTGATCAATTCAAAAGTTGGCGGAACGTTGTGTGTAGATGTAATAGCCAAAAGT 744
QY 165 AspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHis 184
Db 745 GAGGTTCTGATCCAGCTGCTGTATTGATGATGCGATTATCCCGCTAAAAACAAAGGT 804
QY 185 LeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsn 204
Db 805 GTTGATATTATTAATCTGTGTATCCGCTGGACGCTTACAAAATAAAACAAATCTAATGCAA 864
QY 205 GluLeuLysGluValLysGluIleAlaLys-----ProAsnGluIle 218
Db 865 GAATTAGAAAAGTAAAGCGTGAATTAATCGAGCAGTGCAGATGCGGCTCATGAAGCA 924
QY 219 MetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAsp 238
Db 925 TTACTATGTTAGATGCTACAACTGCTCAGATGCTGATGCTCACAAGCTAGAACTTTAA 984
QY 239 AspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGly 258
Db 985 GAAGTAACAAATGTTACAGGTATTGTATTACGAAATTAGATGTCACAGCCAAAGTGCT 1044
QY 259 AlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlu 278
Db 1045 ATCGTATTAGCCATTCGTAATGAATTGCACATTCACAGTTAAATATGATGTTTAGTGAG 1104
QY 279 LysLeuAspGlyLeuGluLeuPheHisProGluArg-----MetAlaSerArg 294
Db 1105 CAATTAGATGACTTACACCATTTAACCTGAAAGTTATGTCACGGCTTATTCGCTGAT 1164
QY 295 IleLeuGlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnAspValAspGln 314
Db 1165 ATGATTGAACAAATATGACGAATAACACAGTTGAAAT---GATCAAAATTGTACAGAA 1221
QY 315 GluLys 316
Db 1222 GAAAG 1227
```

Search completed: February 25, 2003, 03:14:46  
Job time : 109 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 02:59:07 ; Search time 89 Seconds  
(without alignments)  
2871.303 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274  
Sequence: 1 MAFGLSERLQATQMKRGK.....GKGGKRNQNMMLKGNLPPF 455

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool/US09943108/runat\_21022003\_151633\_1953/app\_query.fasta\_1.647  
-DB=Published Applications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZB=500 -MINLEN=0  
-MAXLEN=2000000000 -USBR=US09943108 -CGN\_1\_1\_80\_rnunt\_21022003\_151633\_1953  
-ICPU=6 -ICPU=3 -NO\_XLPHY -NO\_MMAL -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications\_NA:\*  
1: /cgn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2.6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2.6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2.6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2.6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2.6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2.6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2.6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2.6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2.6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2.6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2.6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2.6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2.6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2261	99.4	1368	10	US-09-815-242-8968 Sequence 8968, Ap
2	2235	98.3	1365	10	US-09-815-242-4365 Sequence 4365, Ap
3	2235	98.3	1368	10	US-09-815-242-8030 Sequence 8030, Ap
4	1418.5	62.4	1416	10	US-09-815-242-3890 Sequence 3890, Ap

5	1418.5	62.4	1434	10	US-09-815-242-6764 Sequence 6764, Ap
6	1417.5	62.3	6729	10	US-09-070-927A-37 Sequence 37, Appl
7	1402	61.7	1136	10	US-09-974-300-1847 Sequence 1847, Ap
8	1307	57.5	1572	10	US-09-815-242-9156 Sequence 9156, Ap
9	1306	57.4	1395	10	US-09-815-242-9427 Sequence 9427, Ap
10	1297	57.0	1569	10	US-09-814-041A-1 Sequence 1, Appli
11	1168	51.4	1641	9	US-09-738-626-2255 Sequence 2255, Ap
12	1164.5	51.2	1362	10	US-09-815-242-6166 Sequence 6166, Ap
13	1158	50.9	721	10	US-09-815-242-1949 Sequence 1949, Ap
14	1150	50.6	1374	10	US-09-815-242-7850 Sequence 7850, Ap
15	1138	50.0	1389	10	US-09-815-242-6868 Sequence 6868, Ap
16	1133.5	49.8	1506	10	US-09-815-242-9991 Sequence 9991, Ap
17	1006	44.2	640681	10	US-09-790-988-1 Sequence 1, Appli
18	854	37.6	1347	10	US-09-815-242-7465 Sequence 7465, Ap
19	847	37.2	521	10	US-09-815-242-1985 Sequence 1985, Ap
20	847	37.2	521	10	US-09-815-242-2013 Sequence 2013, Ap
21	764	33.6	468	10	US-09-815-242-3070 Sequence 3070, Ap
22	560	24.6	1440	9	US-09-938-842A-877 Sequence 877, App
23	558	24.5	330	10	US-09-815-242-2247 Sequence 2247, Ap
24	432	19.0	1494	10	US-09-815-242-6281 Sequence 6281, Ap
25	426	18.7	292	10	US-09-815-242-3369 Sequence 3369, Ap
26	418	18.4	1251	10	US-09-815-242-8028 Sequence 8028, Ap
27	416	18.3	948	10	US-09-815-242-4244 Sequence 4244, Ap
28	410	18.0	987	10	US-09-974-300-1852 Sequence 1852, Ap
29	406	17.9	15363	10	US-09-070-927A-279 Sequence 279, App
30	404.5	17.8	1329	10	US-09-815-242-6807 Sequence 6807, Ap
31	404	17.8	1368	10	US-09-815-242-7684 Sequence 7684, Ap
32	403.5	17.7	939	10	US-09-815-242-3937 Sequence 3937, Ap
33	397.5	17.5	1278	10	US-09-815-242-9165 Sequence 9165, Ap
34	397.5	17.5	1293	10	US-09-827-663-1 Sequence 1, Appli
35	396.5	17.4	1290	10	US-09-815-242-9491 Sequence 9491, Ap
36	395.5	17.4	1245	10	US-09-815-242-6997 Sequence 6997, Ap
37	390	17.2	831	10	US-09-827-663-3 Sequence 3, Appli
38	376	16.5	1530	9	US-09-738-626-2259 Sequence 2259, Ap
39	276	12.1	643	10	US-09-815-242-2506 Sequence 2506, Ap
40	268.5	11.8	610	9	US-09-736-457-1094 Sequence 1094, Ap
41	268.5	11.8	610	9	US-09-902-941-1094 Sequence 1094, Ap
42	268.5	11.8	610	9	US-09-849-626-1094 Sequence 1094, Ap
43	267.5	11.8	29793	10	US-09-973-451-38 Sequence 38, Appl
44	193	8.5	405	10	US-09-960-352-13494 Sequence 13494, A
45	189	8.3	420	10	US-09-960-352-14727 Sequence 14727, A

ALIGNMENTS

RESULT 1  
US-09-815-242-8968  
; Sequence 8968, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8968  
; LENGTH: 1368  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1368)  
US-09-815-242-8968

## Alignment Scores:

Pred. No.: 5,28e-245 Length: 1368  
Score: 2261.00 Matches: 452  
Percent Similarity: 99.78% Conservative: 2  
Best Local Similarity: 99.34% Mismatches: 1  
Query Match: 99.43% Indels: 0  
DB: 10 Gaps: 0

US-09-943-108a-2 (1-455) x US-09-815-242-8968 (1-1368)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
Db 1 ATGGCATTGAGGGTATACAGACCCCTTGCAGCGCAGCATGCCAAAATGGTGGTAA 60  
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40  
Db 61 GGTAAACTTACTGAAGCTGATATAAAGATAATCATGCGTGAAGTAGATTAGCGTTACTT 120  
Qy 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60  
Db 121 GAGCGTCAGGTAACCTTTAAAGTGGTAAAGAAATTTATTAACAGTATCAGAACCGCA 180  
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
Db 181 TTAGGTTCCGATGTAATGCATCATTAACACCAGGCAACAAAGTTATTAATAGTCAA 240  
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
Db 241 GATGATTAACCCAGTTCATGGGTGGAGAAAATAGTCGATTAATATGCAAAATAAACCA 300  
Qy 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
Db 301 CCTACTGTTTATGATGGTTGGTTTACAGGTGCTGTGTAACACACACTGCAGGTAAA 360  
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
Db 361 TTAGCATTATTGATCGGTAAATAATACAAAAAACCTATGTTAGTTCAGCAGCATATT 420  
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
Db 421 TATCGTCCAGCAGCGATAATCAATTAACAAACAGTAGGGAACAAATGTATATCTCTGA 480  
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATGATGATTAACATGCT 540  
Qy 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
Db 541 AAAGAAGAACAATTTAGACTTTGTAATCAATTGATACAGCAGGTGCTGATTAACATCGATGAA 600  
Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
Db 601 GCATGATGAACGAATTAAGAAGTAAAGAGCATTTGCTTAACCAACCAAGAAATTTATGTTA 660  
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
Db 661 GTTGTGATTCATCAATGACGGTCAAGATGCTGTCAATGTTGAGAAATCTTTTGACGATCA 720

Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
Db 721 CTTGATGTCACAGGTGTTACCTTAACATAATAGATGGTGATACACGTGGTGCACCT 780  
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
Db 781 TTATCTATTCTGTCGTGACACAAAAACAATTAATTTGTTGGTATGAGTGAAGAAGTTA 840  
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
Db 841 GATGGTTTAGAGCTATTCCATCTGAACGTATGCGATCGATATTTTAGGTATGGGTGAT 900  
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
Db 901 GTGTTAAGCTTAATTTGAAGAAGCGCACACAGATGTGGATCAAGAAAAACAAAGATTTA 960  
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340  
Db 961 GAGAAAAAGATCGGTGATCATCGTTTACTTTAGTATGATTTTATAGACACACTTGCATCAG 1020  
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
Db 1021 GTGAAAAATCTAGGACCCACCTGGATGATATTAATAATGATCCAGGTATGATAAAAATG 1080  
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
Db 1081 AAAGGCTAGATAAGCTTAATATGATGTAAGAACAAATTCATCATTTAAACGATTAATC 1140  
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
Db 1141 CAGTCAATGACGCGCGGTGAAGAAAAAATCCAGACACATTCATGATGATGATGATGATAAAG 1200  
Qy 401 ArgIleAlaLysLysSerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420  
Db 1201 CGTATGCTAAAGGTGCTGCTCTTCATTTACAGAGCTCAATCGTTTGTGATGAACAATTT 1260  
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys 440  
Db 1261 AACGATATGAAGAAATGATGAACACATTCACCTGCGGTGCGGTAAAGTAAAAAGGTAAA 1320  
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
Db 1321 CGCAATCAATGCAAAATATGTTAAAAAGGTATGATTAACGGTTT 1365

## RESULT 2

US-09-815-242-4365  
; Sequence 4365, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4365  
; LENGTH: 1365  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-4365

Alignment Scores: 4, 42e-242 Length: 1365  
Pred. No.: 2235.00 Matches: 447  
Score: 98.90% Conservative: 3  
Percent Similarity: 98.24% Mismatches: 5  
Best Local Similarity: 98.28% Indels: 0  
Query Match: 10 Gaps: 0  
DB:

US-09-943-108A-2 (1-455) x US-09-815-242-4365 (1-1365)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
Db 1 ATGGCAATTCAGAGCTTATCAGAACGCCCTGCAGAGCGACGATGCAAAAAAATCGGTGTAAG 60  
QY 21 GlyLysLeuThrGluAlaAspIleLysLysMetMetArgGluValArgLeuAlaLeuPhe 40  
Db 61 GGTAAACTTACTAGCTGATATAAAGATATATGCGTGAAGTAAAGATTAGCGTTACTT 120  
QY 41 GluAlaAspValAsnPhelLysValValLysGluPheIleLysThrValSerGluArgAla 60  
Db 121 GAGCGTCAGCTAACTTTAAAGTGTAAAGAAATTTATTAACACAGTATCAGACGGCA 180  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
Db 181 TTAGGTTCCGATGTAATGCAATCATTAAACACAGGCAACAAAGTTATTAAATAAGTTCAA 240  
QY 81 AspGlnLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100  
Db 241 GATGAATTAACGAAGTGTAGTGGTGAGAAATAATATPCGATTAATATGTCAAATAACCA 300  
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
Db 301 CCTACTGTTGTTATGATGTTGTTTACAGGTGCTGGTAAACACACACTGCAGGTAAA 360  
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140  
Db 361 TTAGCATTTATGATGCGTAAAAAATACACAAAAAACCTATGTTAGTTGCAGCAGATATT 420  
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160  
Db 421 TATGCTCCAGCAGGATTAATCAATACAAACAGTAGGNAACAATTAATGATTCCTGTA 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
Db 481 TACAGTGAAGAGATCAAGTAAAGCCACACAAATTTGTAACATTAATGATTAACATGCT 540  
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200  
Db 541 AAAGAGAACAATTTAGACTTTGTAATTCATTCATACAGCAGGTGATACACATCGATGAA 600  
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220  
Db 601 GCAATGATGATGAATTAAGAGTAAAGAGATGCTGCTAAACCAACCAAAATATGTTA 660  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240  
Db 661 GTTGTCGATTCATAGCGGCTCAGATGCTGTCATGTTGCGAAGATCITTTGACGATCAA 720  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
Db 721 CTTGATGTCACAGGTGTTACCTTAACCTAAATAGATGGTGATACACAGTGGTGGCAGCT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
Db 761 TTATCTATTCGTTCCGTGACACAAAAACCAATTAATTTGTTGGTATGATGAGTGAAGTTA 840  
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
Db 841 GATGTTTAGAGCTATTCACCTCGAACGTATGGCATCAGCTATTTCCAGGCTGGGTGAT 900  
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320  
Db 901 GTGTCAAGGTTTATTAAGAAAGCGCAACAGATGTGATCAAGAAAAAGCAAAAGATTTA 960  
QY 321 GlnLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
Db 961 GAGAAAAAGATGCGGAGTCATCATTTACTTTAGATGATTTTTTAGACAACACTTGATCAG 1020  
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
Db 1021 GTGAAAAAATTTAGGACCACTGGATGATATTATAAATGATTCAGGATGATGAATAAATG 1080  
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
Db 1081 AAAGGCTTAGATAGCTTAATATAGTGAAGAAACAAATGATCATATTAAAGCGATTATC 1140  
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400  
Db 1141 CAGTCATGACGCGCGCTGAAGAAACAATCCAGACACATTCGATGATCAGTAAAAAG 1200  
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
Db 1201 CGATTGCTAAAGGTTCTGGTTCATTACAGAAGTCAATCGTTTGATGAACAATTT 1260  
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLysLysLysLys 440  
Db 1261 AACGATATGAAGAAAAATGATGAACAATTCATCGTGGCGGTAAGGTAAGGTAAGGTA 1320  
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455  
Db 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGATTTACCGTTT 1365

RESULT 3

US-09-815-242-8030  
; Sequence 8030, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Onlisen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8030
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8030
```

```
Alignment Scores:
Pred. No.: 4,44e-242 Length: 1368
Score: 2235.00 Matches: 447
Percent Similarity: 98.90% Conservative: 3
Best Local Similarity: 98.24% Mismatches: 5
Query Match: 98.28% Indels: 0
DB: 10 Gaps: 0

US-09-943-108A-2 (1-455) x US-09-815-242-8030 (1-1368)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrKetGlnLysMetArgGlyLys 20
Db 1 ATGGCATTGAAGCCTTATCGAACCCCTGCAAGCGAGATGCAAAAATGCGTGGTAAG 60

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTACTT 120

QY 41 GluAlaaspValasnPheLysValLysGluPheIleLysThrValSerGluArgala 60
Db 121 GAGGCTGACGTAACTTTAAAGTGTAAAGAAATTTATTAACACAGTATCATCAAGCGCA 180

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 TTAGGTTCCGTGTATGTCATCATTAACACCGCGGCACCAAGTTATTAATATGTTCA 240

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAAGTTGATGGTGGAGAAATATATCGATTATATGTCAAATAACCA 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CTTACTGTGTATGATGGTTGTTTACAAAGTGTGTTAAACCAACCAACTCGAGGTAAA 360

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATATTGATGCGCTAATAAATACACAAAACCTATGTTAGTTCACGACATATT 420

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATAATCAATTACAAACAGTAGGAAACAAATATGATATTCCTGTA 480

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisala 180
Db 481 TACAGTGAAGGAGATCAAGTAAGTAAAGCCACCAAAATGTAATCATGCAATAAAACATGCT 540

QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleaspGlu 200
Db 541 AAAGAAGAACATTTAGACTTTGTAATCATTGATACACAGCGTCGATTACATCGATGAA 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTTGATGAATGAATTAAGAAGTAAAGACATTGCTAAACCAACCAAGAAATATGTTA 660

QY 221 ValValAspSerMetThrGlyGlnAspAlaValasnValAlaGluSerPheaspGln 240
Db 661 GTTGTGATTCATCAAGCGGTCAAGATGCTGCAATGTTGCAGAACTCTTTGACGATCAA 720

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaIle 260
Db 721 CTTGATGTCACAGGTGTTACTTAATAATTAAGTGTGATACACGTGGTGGTGCAGCT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
```

```
Db 781 TTATCTATTCTGTCGGTGACACAAAACCAATAATTTGTTGGTATGAGTGAAAGTTA 840

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetalaSerArgIleLeuGlyMetGlyasp 300
Db 841 GATGTTTTAGAGCTATTCCATCTCTGAACGATATGCGATCAGCATATTTTCAGCCTGGTGTAT 900

QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnaspValaspGlnGluLysAlaLysaspLeu 320
Db 901 GTGTCAGGTTANTTTGAAAAGCGCAACAGATGGGATCAAGAAAAGCAAAAGATTATTA 960

QY 321 GluLysLysMetArgGluSerSerPheThrLeuaspaspPheLeuGluGlnLeuaspGln 340
Db 961 GAGAAAAGATGCGCGAGTCATCATTTACTTTAGATGATTTTITAGACAACTTGATCAG 1020

QY 341 ValLysAsnLeuGlyProLeuaspaspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATTTAGGACCACTGGATGATATTATTAATAATGATCCAGGTATGATAAAATG 1080

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGTCTAGATTAAGCTTAATATGAGTGAAAAGCAAAATGATCATATTAAAGCGATTATC 1140

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProaspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACGCCGCTGAAAGAAACAATCCAGACACATTGAATGATATATCAGTAAAAAG 1200

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValasnargLeuMetLysGlnPhe 420
Db 1201 CGTATTGTAAAGTTCTGCTGCTTCATTACAAAGAGTCAATCGTTGATGAACAATTT 1260

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
Db 1261 AACGATATGAGAAAATGATGAAACAATTCATCTGCTGCGGTAAAGTAAAGGTTAAAG 1320

QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProphe 455
Db 1321 CGCAATCAAAATGCAAAATATGTTAAAAAGGTATGAATTTACGGTTT 1365
```

```
RESULT 4
US-09-815-242-3890
; Sequence 3890, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 3890
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3890

Alignment Scores:
Pred. No.:      3,17e-150      Length:      1416
Score:          1418.50      Matches:      273
Percent Similarity: 77.54%      Conservative: 86
Best Local Similarity: 58.96%      Mismatches: 89
Query Match:      62.38%      Indels:      15
DB:              10          Gaps:         2

US-09-943-108a-2 (1-455) x US-09-815-242-3890 (1-1416)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCTTTTGAGAGTTTAAACAAACCGCTACAAACAGGCAATGAGTAAATCCGTCGTAAG 60
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAAGAGTTCCGAGCGCGAGCTAAAGAAATGATCGGAGAAATCCGTTGGCTTTATTA 120
Qy 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAAGCGAGCTTAATTTACAAGTGTCAAGATTTTCACAAACGCGTCAGAGAACGGCA 180
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTAGGAGTGAAGTATTAGAAGCTTATCACCAGCCCAACAAATTTGAAAAATTTGTGAT 240
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAAGAATTACGAACAACTTAGGTCAGAAACCGTTGAACCTGAATTAATCTCCAAAATC 300
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CGACAGTGATGATGACAGGTTTACAAGGGCTGTGTAACAACTTTTACTGGTAA 360
Qy 121 LeuAlaLeuLeuMetArgLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTACAAAACACTTAATGAAAACCTGAAACGCTCGTCCGCTTTAATCGCTGGTGAAGTT 420
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATGATGATGAGGTTTAGGTCACAACTTAGAAGTTCCCGTT 480
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTTGATATGGACAGACAGCTAATCCAGTGGAAATTTGTCGCAAGGCTTAGCATTAGCA 540
Qy 191 LysGluGlnHisLeuAspPheValIleAlaAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAAAGAAAATGANTGTCTTAATTGATACGCGCGCGCTTACACATTGACGAA 600
Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCITTAATGACCAATGAAACAAATTAAGAGTTGGCTTAATCCCAATGAAATTTCTGTTA 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATCGGATGACGGGCAAGATGCTCTCAAGTTGCGAGATAGTTTAAATGACAC 720
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGGAATTACTGGGTTCTTATTACCAAAAATGACGGCGGATACTCGTGGGGGGCGTGC 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGCAATTGGGCGATACGGGCGCTCCGATTAAATTTGTCCGTTCTGGTGAAAAATTA 840
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
```

```
RESULT 5
US-09-815-242-6764
; Sequence 6764, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIORITY FILING DATE: 2001-03-21
; PRIORITY FILING DATE: 2001-03-21
; PRIORITY FILING DATE: 2000-03-21
; PRIORITY FILING DATE: 2000-03-21
; PRIORITY FILING DATE: 2000-05-23
; PRIORITY FILING DATE: 2000-05-23
; PRIORITY FILING DATE: 2000-05-26
; PRIORITY FILING DATE: 2000-05-26
; PRIORITY FILING DATE: 2000-10-23
; PRIORITY FILING DATE: 2000-10-23
; PRIORITY FILING DATE: 2000-11-27
; PRIORITY FILING DATE: 2000-11-27
; PRIORITY FILING DATE: 2000-12-22
; PRIORITY FILING DATE: 2000-12-22
; PRIORITY FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 6764
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
US-09-815-242-6764

Alignment Scores:
Pred. No.: 3,23e-150 Length: 1434
Score: 1418.50 Matches: 273
Percent Similarity: 77.54% Conservative: 86
Best Local Similarity: 58.96% Mismatches: 89
Query Match: 62.38% Indels: 15
DB: 2

US-09-943-108A-2 (1-455) x US-09-815-242-6764 (1-1434)
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB 1 ATGGCTTTTGAGAGTTTAAACACCCCTACACAGGAGTGAAGTAAATCCGTCGTAA 60

QY 21 GlyLysLeuThrGluAlaAspLysLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
DB 61 GGAAGAAGTTCGGAAGCGGAGTAAAGAAATGATGCGAGAAATCCGTTGGCTTTATTA 120

QY 41 GluAlaAspValAsnPheLysValLysGluPheLeuLysThrValSerGluArgAla 60
DB 121 GAAGCGGAGTTTAAACAGGTTTACAGAGTGGTCAAGAGTTTACAAACCGCTCAGACAGCGGCA 180

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLeuValGln 80
DB 181 GTAGGAGTCGAAGTATTAGAAAGCTTATCACCAGGCCAACAAATGTAATAATTTCTGAT 240

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysMetSerAsnLysPro 100
DB 241 GAAGAATTAACGAACAAACGTTAGTGTGAGAAACGTTGAACTGAATAAATCCCAAAATC 300

QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
DB 301 CCGACAGTGATATGATGACAGGGTTTACAGGGGCTGTGTAACAACTTTTACTGGTAA 360

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspLe 140
DB 361 TTAGCAAAACACTTAATGAACAACTGAAACCGCTCGCTGCTTTTAAATCGCTGTGACGCT 420

QY 141 TyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLysLeuAspLysProVal 160
DB 421 TATCGTCCAGCAGCGATGTATGATGATGAGAGTTTATGAGTCAACATTAAGAGTCCCGTT 480

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180
DB 481 TTTGATATGGGAACAGATGCTAATCAGTGGAAATTTGTCTCAAGGGTTAGCATTAAGCA 540

QY 181 LysGluGluHisLeuAspPheValLysLeuAspThrAlaGlyArgLeuHisLysLeuAspGlu 200
DB 541 AAGAAGAAGAAATGATATGCTTAAATGATAGCGCGCGCTTTTACACATTAAGCAAG 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLeuLysProAsnGluLysLeuMetLeu 220
DB 601 GCCTTAATGGAGCAATTAACAAACAAATTAAGAGTTGGCTTAATCCCAATGAATTTCTGTA 660

QY 221 ValValAspSerMetThrGlnGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
DB 661 GTTGTGTATGCGATGACGGGCAAGATGCTGTCACCTGTGACAGATGTTTAAATGAACAG 720

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
DB 721 CTGGAATTTACTGGGTTGTTTATTAACAAATGAAGCAGCGGTAATCTCTGGGGGGCTGCG 780

QY 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280

```



APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6729 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 37:  
US-09-070-927A-37

Alignment Scores:  
Pred. No.: 3,83e-149 Length: 6729  
Score: 1417.50 Matches: 273  
Percent Similarity: 77.54% Conservative: 86  
Best Local Similarity: 58.96% Mismatches: 89  
Query Match: 62.34% Indels: 15  
DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-070-927A-37 (1-6729)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
|||||  
Db 3722 ATGGCTTTGAGAGTTTAAACAAACCGCTACAGCAGGCAATGAGTAAATCCGCTGTAAG 3781  
|||||

Qy 21 GlyLysLeuThrGluAlaAspIleGlyLysMetMetArgGluValArgLeuAlaLeuPhe 40  
|||||  
Db 3782 GGAAGAGTTCCGAGACCGCGCTAAAGAAATGATCGAGAAATCCGTTGGCTTATTA 3841  
|||||

Qy 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60  
|||||  
Db 3842 GAAGCCGAGCTTAATTTACAGTGGTCAAGAGATTTCCACAAACGCTGTCAGAGACGGCA 3901  
|||||

Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80  
|||||  
Db 3902 GTAGGAGTCAAGATTAAGAAAGCTTATCACCAGCCCAACAAATTTGTAATAATTTGAT 3961  
|||||

Qy 81 AspGluLeuThrLysLeuMetClyGlyLysThrSerIleAsnMetSerAsnLysPro 100  
|||||  
Db 3962 GAAGAATTAACGAAACACTAGGTTCAGAAACCGTTGAACTGAATAATCTCCAAATC 4021  
|||||

Qy 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120  
|||||  
Db 4022 CCGACAGTGATGATGACAGGCTTACAGGGGCTGTTAAACAACTTTTACTGTGATAA 4081  
|||||

Qy 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspIle 140  
|||||  
Db 4082 TTACAAACAACTTAATGAAACACTGAAACCGCTCGTCCGCTTTTAAATCGCTGGTACGT 4141  
|||||

Qy 141 TyrArgProAlaAlaIleAsnGlnGluThrValGlyLysGlnIleAspIleProVal 160  
|||||  
Db 4142 TATCGTCCAGCAGCGATGATGATGAGTGGTGAAGTTCAGTCAACATTTAGAGTTCCTGTT 4201  
|||||

Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180  
|||||  
Db 4202 TTGATATGGGACAGATGCTAATCCAGTGAATTTGTTCTGTTCAAGGTTAGCATTAAGCA 4261  
|||||

Qy 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200  
|||||  
Db 4262 AAAGAAAGAAATGATGATGCTTAAATGATGATGATGATGATGATGATGATGATGATGAT 4321  
|||||

Qy 201 AlaLeuMetAsnGluLeuLysGlnValLysGluIleAlaLysProAsnGluIleMetLeu 220  
|||||  
Db 4322 GCTTAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4381  
|||||

Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240  
|||||

Db 4382 GTTGTGTGTCGATGACGGGCGAGATGCTGTGCACGTTGTCAGATAGTTTATGACAG 4441  
|||||

Qy 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260  
|||||

Db 4442 CTGGAATTAATCTGCGGTTGTTATTAACAAATGATGCGGATGCTGCTGGGGGGTGG 4501  
|||||

Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280  
|||||

Db 4502 CTGTCAATTCGGGCGAGTAACGGGTGCTCCGATTAATTTGCGGTTCTGTTGAAATTA 4561  
|||||

Qy 281 AspGlyLeuGluPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300  
|||||

Db 4562 ACCGATTTAGAAATTTCCATCCCGATGCTGATGCGAGTCGATATGATGATGATGATGAT 4621  
|||||

Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320  
|||||

Db 4622 ATGTTGACGCTAATTTGAAAAAGCCACACAGATTACGATGAGAAAAAGCAGAACTT 4681  
|||||

Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
|||||

Db 4682 GCTCAAAAAATGAAAGAAACAGTTTTCGCTTAACGATTTTCATTTGACCAATTTGATCA 4741  
|||||

Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360  
|||||

Db 4742 GTTATGGCATGGGACCGATGAGACTTATTAATAATGATCCCTGGAATGAGTAACATG 4801  
|||||

Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380  
|||||

Db 4802 CTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4861  
|||||

Qy 381 GluSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400  
|||||

Db 4862 TTATCAATGACCCCTGCGAGACGTTGAAATCCCTGATCTATTAAATCTCTAGTCGCCGCC 4921  
|||||

Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420  
|||||

Db 4922 AGAATTCGACGCTGTTGAGGAAATGATGTTGTTGAGTCAATCGTATGATTAACAATTT 4981  
|||||

Qy 421 AsnAspMetLysLysMetLysGlnPheThr----- 431  
|||||

Db 4982 AAAGAAATCCAAAAATGATCAACAAATGTCAAAAGGATGTCGACGAAACGGCGATGGTC 4861  
|||||

Qy 432 -----GlyGlyGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLys 5041  
|||||

Db 5042 GATCAATGTTAGTGGCGGCTTAAAGGCAAGCTAGTAAATG---GCCATGATCGT 5098  
|||||

Qy 447 MetLeuLys 449  
|||||

Db 5099 ATGATGAAG 5107  
|||||

RESULT 7  
US-09-974-300-1847  
; Sequence 1847, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: 10085:500-US  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1847  
; LENGTH: 1136  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis

US-09-974-300-1847

## Alignment Scores:

Pred. No.: 1,66e-148 Length: 1136  
Score: 1402.00 Matches: 270  
Percent Similarity: 87.30% Conservative: 60  
Best Local Similarity: 71.43% Mismatches: 48  
Query Match: 61.65% Indels: 0  
DB: 10 Gaps: 0

US-09-943-108a-2 (1-455) x US-09-974-300-1847 (1-1136)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
DB 1 ATGCAATTCGAGGATTAGCGCGAGTGCAGCAAAACGATCTCAAAATCCGCGGAAA 60  
QY 21 GlyLysLeuThrGluAlaAspLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40  
DB 61 GGAAGGTAAACGAGCAAGACGCTCAAGAGATGATGCGGAAATCGGTTAGCGCTGCTT 120  
QY 41 GluAlaAspValAsnPhelLysValLysGluPheLeLysThrValSerGluArgAla 60  
DB 121 GAACGGAGCTCAATTTAACTGCTGAAGATTTGTGAAAAGGTAAAGCGAGCT 180  
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLeuValGln 80  
DB 181 CTCGGCAAGACGCTTATGAAAAGCTGACGCCGACAGCATCAATTAAGTGTAAA 240  
QY 81 AspGluLeuThrLysMetGlyGlyGluAsnThrSerLeuAsnMetSerAsnLysPro 100  
DB 241 GAAGAGTGCAGGAGTGTGGCGGCGGAGAAAGCAAGATTGCCGCTGCAAGCGTCCG 300  
QY 101 ProThrValValMetValGlyLeuGlnGlyValGlyLysThrThrAlaGlyLys 120  
DB 301 CCGACCGTCATCATGATGTCGCTGTCAGGGCGCGTAAACAGCATGTCACCGGAAAG 360  
QY 121 LeuAlaLeuLeuMetArgLysLysTyraAsnLysLysProMetLeuValAlaAlaAspLe 140  
DB 361 CTTCGCAACGCTGCTCGCAAAACATAAACGCAATCCGCTGCTGTCGACGACATC 420  
QY 141 TyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLysLeuAspLeuVal 160  
DB 421 TACCGCCGCGCGCATCAACAGCTGGAACGCTCGGCAACGCTCGATATCGCGGTA 480  
QY 161 TyrSerGluGlyAspGlnValLysProGlnLysLeuValThrAsnAlaLeuLysHisAla 180  
DB 481 TTTTCATCGGTGACAGTACGTCGCTGAGATGCCAGACGATCGCAAAAGCG 540  
QY 181 LysGluGluHisLeuAspPheValLeuLeuAspThrAlaGlyArgLeuHisLeuAspGlu 200  
DB 541 AAAGAAGATCATCAGCATACGCTTCTCATCATACGCGGAGCGCTTCATATTGATGA 600  
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeu 220  
DB 601 GAGTGTATGATGAGTTCAGCAGTAAAGAAACCGCTCAGCGGAAAGAGATTTTCCTT 660  
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAlaAsnValAlaGluSerPheAspAspGln 240  
DB 661 GTCGTTGACTCCATGACGGGTCAACCGCGTCAATGTTGCCAAAGCTTTAAGCAACAG 720  
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260  
DB 721 CTCGCTGTGACAGGTGTCATCTTGACGAAGCTTGACGGCGATAGAGCGGAGCGGCC 780  
QY 261 LeuSerLeuArgSerValThrGlnLysProLeuLysPheValGlyMetSerGluLysLeu 280  
DB 781 TTGTCAATCCGCGCGCTGCAACACCGCGATCAATTCGCGGATGCGGGAAGGCTT 840  
QY 281 AspGlyLeuGlnLeuPheHisProGluArgMetAlaSerArgLeuGlyMetGlyAsp 300  
DB 841 GATGGCTGGAGCGGCTTTCATCTCAACGAGTGGCTTCGCGGATATCTCGAATGGCGAT 900  
QY 301 ValLeuSerLeuLeuGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320

DB 901 GTGCTACCCCTTATCGAGAAAGCCAGGCCAACGTCAGCAAGAAAGCGAAGAACTC 960  
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340  
DB 961 GAGCAAAATGAAATATGAGCTTTACGCTGACGACTTTTGGAAACAGCTCGGACAG 1020  
QY 341 ValLysAsnLeuGlyProLeuAspLysMetLysMetLysProGlyMetAsnLysMet 360  
DB 1021 GTCCGCAATATGCGCGCTTGAGAGCTGATCAATGATGCCGCGGACGCAAAATG 1080  
QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnLysAspHisLeuLysAla 378  
DB 1081 AAAGCTTAAATAAACGTCGAAGCTTGATTAATAAACAGCTCAGCATATCGAAGCG 1134

## RESULT 8

US-09-815-242-9156  
; Sequence 9156, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9156  
; LENGTH: 1572  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1572)  
US-09-815-242-9156

Alignment Scores:  
Pred. No.: 1,28e-137 Length: 1572  
Score: 1307.00 Matches: 253  
Percent Similarity: 74.73% Conservative: 90  
Best Local Similarity: 55.12% Mismatches: 102  
Query Match: 57.48% Indels: 14  
DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-815-242-9156 (1-1572)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20  
DB 1 ATGCAATTCGAGGATTAGCGCGAGTGCAGCAAAACGATCTCAAAATCCGCGGAAA 60  
QY 21 GlyLysLeuThrGluAlaAspLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40

```
Db 61 GGAANAATCTGTAATCTGATGTCORAGAGGCAACCAAGAAATTCGCTTGCCCTGCTC 120
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 41 GluAlaAspValAsnPhelYsValValYsGluPhelLeYsThrValSerGluArgAla 60
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 121 GAGGCGAGCTTGCTTGCTGCTGTTGTAAGAGGCTTTATCAAGAAAGTTCTGAGAGTGCA 180
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 181 GTCGGGCATGAGGTCATTGATCACTTAATCCTCGCAACAGATATTATAAATCGTTGAT 240
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 241 GAGGAAGTACAGCCGTTTATGCTGATACGCGAGAAATATCAAGTCACTCAAGTAT 300
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 301 CCAACCATCATCATGATGTTGTTTACAAGGGGCTGTAACAAACATTTGCTGTGTA 360
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 361 TTGGCCCAACAACCTCAGAAGAAGAAATGCTGCTCTTTGATGGTTGGCGGGATAT 420
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 421 TATCGTCCAGCTGCCATTCACCACTTAAGACCTTGGGACACAGATGATGTCGCTGC 480
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 481 TTTGCACCTTGAAGCAAGAACTACCAAGCTGTGTGATGATTCACGTCAAGGTTTGGAGCAAGCC 540
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 541 CAACATTAATCAATACGACTATGCTGATGATGACTGCGGGTTCGATTCGACATGATGAG 600
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 201 AlaLeuMetAsnGluLeuLysGluValIleAlaLysProAsnGluIleMetLeu 220
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 601 CTCCTCATGATGACCTTCGTGATGTGAAGCATTTGGCTCAACCAATGAATCTTGCT 660
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 661 GTGGTGTGATGCTATGATGTTGTCAGGAAGCAGCAATGTTGGCGCTGATGTTAATGCTCAG 720
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 721 TTGAAGTACGCTGGGCTCATCTTACCAGATGATGGCGATPACTCGTGGTGGTGTGCT 780
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 781 CTGCTGCTGCTGACATCTACCTGGAACCAATCAAGTTCACGTGCTACAGGTGAAGAAAT 840
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgTleLeuGlyMetGlyAsp 300
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 841 ACGGACATTTGAACACCTTCACCCAGACCGCATGCTAGCCGTATCCTGTTGTTATGGGGAT 900
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 301 ValLeuSerLeuIleLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 901 ATGCTCATCTTGTATGAGAAGCTTCTCAGGAATACGATGAACAAAAGCCCTTGAATG 960
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 321 GluLysLysMetArgLysSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 961 GCTGAGAAGATGCGGCAACACCTTTGATTAATGATTTTCATCGATCAATAGATCAG 1020
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 341 ValLysAsnLeuGlyProLeuAspIleMetLysMetIleProGlyMetAsnLysMet 360
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 1021 GTGCAAAATATGGGCGGATGGAAGACTTGTCTCAAGATGATTCACAGTATGGCCCAACAAT 1080
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleAlaIleIle 380
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 1081 CCAGCCCTTCAAAACATGATGAGGTGATGAACGCCGATTCGTCGTAACGTCGCAATG 1140
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
```

```
Db 1141 TCTTCGATGACACCTGAAGAGCGTGAAGACCCAGATTTGTTAAATCCAGCCGCTCGCCGT 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 1201 CGTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 1261 AACCAAGGCTAACACCTCATGACGGGTGTTATGCTCTGGG----- 1299
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 1300 -----GATATGAATAAATGATGAACCAAAATGGGATTAATCAAAATAACCTTCT 1350
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

RESULT 9
US-09-815-242-9427
; Sequence 9427, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9427
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1395)
US-09-815-242-9427

Alignment Scores:
Pred. No.: 1.39e-137 Length: 1395
Score: 1306.00 Matches: 252
Percent Similarity: 74.73% Conservative: 91
Best Local Similarity: 54.90% Mismatches: 102
Query Match: 57.43% Indels: 14
DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-815-242-9427 (1-1395)
Qy 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 1 ATGCATTTCAAAAGTTTACAGACGTTTGCAGACGCTTTTAAATAATCTACGTAATAA 60
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgLysValArgLeuAlaLeuPhe 40
      |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
```

Db 61 GGAARAATCTCGAATCTGATGTCCAAAGAGCAACCAAGAAATTCGGCTGGCTGC 120

Qy 41 GlualaaspValasnPhelysValValylsclupheilelysthrValSerGluA-gala 60

Db 121 GAGGCGAGCTTGCGCTTGCTGTTGAAGAGCTTTATCAAGAAAGTTTCGTCGTCGA 180

Qy 61 LeuglySeraspValMetGlnSerLeuthrProGlyGlnValIlelylValGln 80

Db 181 GTGGCGATGAGTCATTCATGATCACTTAATCCTCGCAACAGATTATTAATTCGTTGAT 240

Qy 81 AspGluLeuthrLysLeuMetGlyGluAsnThr-SerIleAsnMetSerAsnLysPro 100

Db 241 GAGGAAGTACAGCCGTTTGTAGTCTGATACCGCAAGAAATTCACAGTACCTAGATT 300

Qy 101 ProthrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120

Db 301 CCAACCATCATCATGATGTTGGTTTACAAGGGCTGGTAAACAACCTTTGCTGGTAAA 360

Qy 121 LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIle 140

Db 361 TTGGCCACAAATCTCAAGAAAGAAATGCTCGCTTGTGATGATTCGGCGGATATT 420

Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160

Db 421 TRTGTCGAGCTGCCATGACCACTTAAGACCTTGGACACAGATTGATGCTCGTGC 480

Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180

Db 481 TTTCGACTTGGACAGAAAGTACCAGCTGTTGAGATTGTACGTCAAGGTTTGGACCAAGCC 540

Qy 181 LysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200

Db 541 CAAACTAATCATCAACAGCATGTCTGATTGATGACTCGCGGTCGTTTCAGATTGATGAG 600

Qy 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220

Db 601 CTCCTCATGATGAGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 660

Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240

Db 661 GTCTGTCATGCTATGATGTTGGTCAGAAAGCAAGCAAGTTCGCGTCGTCGTCGTCGTC 720

Qy 241 LeuAspValThrGlyValThrLeuthrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260

Db 721 TTGAAGTGAAGTGGGTCATCCTACCAAGATTGATGCGGATACCTCGTGGTGGTGTCT 780

Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280

Db 781 CTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 840

Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300

Db 841 ACGGACATTGAAACCTTCACCAGCCGATGCTAGCGGTATCCTGTCGTCGTCGTCGTCG 900

Qy 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320

Db 901 ATGCTCACTTGTATGAGAAGCTTCACGAATACGATGAACAAAGAGCCCTTGAATG 960

Qy 321 GlulysLysMetArgGlnSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340

Db 961 GCTGAGAGATGCGCAACACCTTTCATTTAATGATTCATGATCAATAGATGATGAT 1020

Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360

Db 1021 GTGCAAAATATGGGCGGATGAAGACTTGCTCAAGATGATTCAGGTCATGGCCAAAT 1080

Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380

Db 1081 CCAGCATCTCAAAACATGAAGTGGATGAACCGCAGATTGCTCGTAAACAGTCGTCATTGTG 1140

Qy 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400

Db 1141 TCTTCGATGACCTGAAGACGTCGAACCCAGATTGTTAAATCCAGCCGTCGCGGT 1200

Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420

Db 1201 CGTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260

Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440

Db 1261 AACCAGGCTAAACAGCTCATGCGGGTGTATGCTGGG----- 1299

Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454

Db 1300 -----GATATGAATAAATGATGAGCAAAATGGGATTAATCCAATAACCTTCCT 1350

RESULT 10

US-09-814-041A-1

; Sequence 1, Application US/09814041A

; Patent No. US20020103104A1

GENERAL INFORMATION:

; APPLICANT: CHEEVER, CHRISTY

; APPLICANT: FECTEAU, DOUGLAS A.

; APPLICANT: LI, HU

; APPLICANT: PAYNE, DAVID J.

; APPLICANT: STEEL, ANGELA

; APPLICANT: WANG, LEI

; TITLE OF INVENTION: METHODS USING THE SRP POLYNUCLEOTIDES

; FILE OF INVENTION: AND POLYPEPTIDES COMPOUNDS MODULATING THEIR ACTIVITY

; FILE REFERENCE: GM50069

; CURRENT APPLICATION NUMBER: US/09/814,041A

; PRIOR FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/191,008

; PRIOR FILING DATE: 2000-03-21

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 1569

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-814-041A-1

Alignment Scores:

Pred. No.: 1,7e-136 Length: 1569

Score: 1297.00 Matches: 251

Percent Similarity: 74.51% Conservatives: 91

Best Local Similarity: 54.68% Mismatches: 103

Query Match: 57.04% Indels: 14

DB: 10 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-814-041A-1 (1-1569)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20

Db 1 ATGCGATTTGAAAGTTTAAACAGAGCTTTCAGAGAGCTCTTTTAAATACTACGTAAAAA 60

Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaPhe 40

Db 61 GGAARAATCTCGAATCTGATGTCCAAAGAGCAACCAAGAAATTCGGCTGGCTGC 120

Qy 41 GluAlaAspValAsnPhelysValValylsclupheilelysthrValSerGluArgAla 60

Db 121 GAGCGCAGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180

Qy 61 LeuGlySeraspValMetGlnSerLeuthrProGlyGlnGlnValIlelylValGln 80

Db 181 GTGGCGATGAGTCATTCATGATCACTTAATCCTCGCAACAGATTATTAATTCGTTGAT 240

Qy 81 AspGluLeuthrLysLeuMetGlyGluAsnThr-SerIleAsnMetSerAsnLysPro 100

Db 241 GAGGAATGACAGCCGTTTGTAGTCTGATACCGCAAGAAATTCACAGTACCTAGATT 300

Qy 101 ProthrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120

Db 301 CCAACCATCATCATGATGTTGGTTTACAAGGGCTGGTAAACAACCTTTGCTGGTAAA 360

```
QY 121 LeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTGGCCAAACAACTCAGAAAGAAAGAAATGCTGCTTGAAGATTGCGCGGAAAT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCAGCTGCATTGACAGCTTAAGACCTTGGACACACAGATTGATGTGCTGTC 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTGGCACTTGAACAGAAAGTACAGCTGTGAGATTACGTCAGAGTTGGAGCAAGCC 540
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 CAACTAATCATACAGCATATGCTGTGATTGATCTGCGGGTCTTGGAGATTGATGAG 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 CTCCTCATGAATGAGCTGCTGATGTGAAGTATTGGCTCAACCAAAATGAATCTTGCT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTCGTTGATGCTATGATGTGTAGGAAGCAGCCATGTTCGCGTGAGTTTAATGCTCAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 TTGGAAGTACTGGGGTCATCCTTACCAAGATTGATGCTGATCTACTGCTGGTGTCTCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGTCTGTCTCATCATCTACATGTAACCAATCAAGTCTACTGTCAGAGTGAAGAAAT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACAGATATCGAAACCTTCCACCAGACGATGTCTAGCGATATCCTTGGCATGGGGAG 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 ATGCTCACTTTGATGGAAGCTTCTCAGGAATACGATGACAAAGCCCTTGAATG 960
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAAGATCGCGAAACACCTTTGATTTTAATGATTTTCAATGATCAATATAGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGCAAAATATGGGCGCGATGGAGACTTCTCAAGATGATTCAGATATGCGCCACAAAT 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCATTCAAAACATGAAGTGGATGAACCCAGATTGCTGTAACGTCGCAATTGG 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACATCTGAAGACGCTGAAACCCAGATTTGTTAAATCCAAACCGCTCGCGT 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTGCTGGTTCGGAATATACATTGCTGGAAGTCAATAATATTCATCAAGGACTTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysLysLys 440
Db 1261 AACCAGGCTAAACGCTCATGCGAGGTGTTATGCTGGG----- 1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATGATGAAGCAAAATGGGGATTAAATCCAAATAACCTTCT 1350
RESULT 11
US-09-738-626-2255
; Sequence 2255, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
```

```
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2255
; LENGTH: 1641
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2255

Alignment Scores:
Pred. No.: 5.84e-122 Length: 1641
Score: 1168.00 Matches: 237
Percent Similarity: 66.67% Conservative: 85
Best Local Similarity: 49.07% Mismatches: 125
Query Match: 51.36% Indels: 36
DB: 9 Gaps: 5

US-09-943-108a-2 (1-455) x US-09-738-626-2255 (1-1641)
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTTGAGTCTACTGTCGATCGGTGTAATGAGCGGCTTCCGGCCCTGCGCGCAAGGAAG 53
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
Db 64 CTCACCGGCGAGACATCAATGCAACACACGCGAGATCCGCTCTCGCGCTGCTGGAAGCT 123
QY 43 AspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db 124 GACGTTTCAATAACGGTTGTTCTGCTTCAATAACCGAATCAAGGAACGCGCGCTGTGT 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
Db 184 GCAGAAGTTTCTCAGGCACTCAACCCGCGCGCAGCAAGTCAATCAAGATCGTCAACAGGAA 243
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
Db 244 CTGTTTCAGATCCTCGGTGGCGAAACCCGCGACTGTCTACTGGCCAAAACCCACCGAAC 303
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTCATCATCTCGCAGGTCGACGGGTGCGAGGTGAAGACCCCTCCGAGGTAACCTGTC 363
QY 123 LeuLeuMetArgLysLysTyrAsnLysLysPheMetLeuValAlaAlaAspIleTyrArg 142
Db 364 AACACCTGTCACAGCAG---GGTCACACTCCTATGCTGTGTGCTGTGACCTTCAGCT 420
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyr--- 161
Db 421 CCAGCGCAGTTTCAACGACTGCAATTTGTGGTGAACGCGCAGGCGTTTACCCTTCGCA 480
QY 162 -----SerGluGlyAsp 165
Db 481 CCGGATCCAGGCACCACGATCGACTCCCTCGACGAAATGGGCACTCCCGCGGTGAT 540
```

```
Qy 166 GlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGlnHisLeu 185
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 -----CAGTCGAGGTAGCGCGCAGGTATCGAGAGCCAGCGCACCCAGCAC 591

Qy 186 AspPheValIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGlu 205
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 592 GACATCGTGATCGTGAACCGCAGGTGCGTGGTATCGATGAACCCGTGACTCAG 651

Qy 206 LeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSerMet 225
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 GCACGCAACATCGGAGACCATCAACCCGTGATGAAGTCTCTTTGTCATTGACTCCATG 711

Qy 226 ThrGlnGlnAspAlaValAsnValAlaGlnSerPheAspAspGlnLeuAspValThrGly 245
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 712 ATTGGTCAGAGCGCGTAGACCGCGAGAGCATCCCGCGCGGTGCTGACACTTCCACCGT 771

Qy 246 ValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArgSer 265
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 772 GTTGCTCTCACCAGCTTGATGGCGAGCGCGCGGTGGTGTGCTGACACTATCCATCCGTGAA 831

Qy 266 ValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeu 285
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 832 GTCACGCGCAACCCATCATGTTGGCTCCACTGTGTGAAAACTCGACGACTTCGACGTC 891

Qy 286 PheHisProGlnArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIle 305
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 892 TTCACCCAGAGCGCATGCCAGCGAATCTTGGCATGGGTGAGTACTGTCACTCATC 951

Qy 306 GluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGluLysLysMetArg 325
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 952 GAGCAGCGCGAGCAGTATGATGATCAGAAAAAGCAGAGGTGCTGCCCGCAAGTTGGCG 1011

Qy 326 GluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGly 345
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1012 TCGGGGAGCTCACCTGGAGAGCTTCTTGACCAATGCTGATGATCCGCGCATGGGA 1071

Qy 346 ProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLys 365
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1072 CCAATCGCAACATCCTCAAGATCTGCTGTTGGCAGCAGATGTCCTCAAAATGGCGGAC 1131

Qy 366 LeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleGlnSerMetThrPro 385
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1132 ATG---GTTGATGAGAACCACTCGACCGCATCCAGGCGATATCCGCGTATGACCCCG 1188

Qy 386 AlaGluArgAsnAspProAspThrLeuAsnValSerArgLysLysArgIleAlaLysGly 405
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1189 GCGAGCGCGATATCCAAAGATCCTCAAGCTTCAGCGCTCCAGCGCAAGCGCATCGCCACGCT 1248

Qy 406 SerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPheAsnAspMetLysLys 425
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1249 TCGGGTGTACCGTGTCCGAAAGTAAACAACTTGTTGAACGCTTCTTCGAGGCTCCCAAG 1308

Qy 426 MetMetLysGlnPheThrGly----- 432
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1309 ATGATGGGTCAATCGCTGCGCAGATTTGGCATGGGTCTCGATCCCGCATGCAACCAAG 1368

Qy 433 ---GlyGlyLysGlyLysArgAsnGlnMetGlnAsnMetLeuLysGlyMet 451
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1369 AACCAAGCCCAAGCGCGAGGTGAGACGGCAAGCGTAAACCAAGCAAGAGGGGCCCA 1428

Qy 452 AsnLeuPro 454
|||
Db 1429 ACCAGCCA 1437
```

## RESULT 12

US-09-815-242-6166

; Sequence 6166, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

```
; APPLICANT: Wall, Daniel
; APPLICANT: Travick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6166
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1362)
US-09-815-242-6166

Alignment Scores:
Pred. No.: 1,11e-121 Length: 1362
Score: 1164.50 Matches: 231
Percent Similarity: 71.74% Conservative: 94
Best Local Similarity: 50.99% Mismatches: 121
Query Match: 15.21% Indels: 7
DB: 10 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-815-242-6166 (1-1362)
```

```
Qy 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 TTGTGATAATTTAACCGATGCTTGTGCGCAGCGCTGCGCAATATCAGTGGCCGTGGACGC 63

Qy 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 CTCACGTAGACAGACAGTAAAGATACGCTGCGGGAAGTGCAGTGGCGCTGCGAGGGG 123

Qy 43 AspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 GAGCTAGCTCTCGCGTAGTCTGCTGAGTTTATCAATCGGTAAGAGAGAGAGCGGTGCT 183

Qy 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 CATGAAGTTAATAAGAGGCTCACCGCGGCGAGGAGTTCGTCAAAATAGTCCGTAAAGAA 243

Qy 83 LeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 244 CTGGTTGCGCGCATGGCGAGAGACACACCCCTGACCTGGCTGGCGACCGCTGGG 303

Qy 103 ValValMetMetValGlyLeuGlnGlyLysThrThrAlaGlyLysLeuAla 122
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 GTCGTACTGATGGCGGCTTCAAGGTGCGGTAAACACACACCGCTTGTGAACCTGCT 363

Qy 123 LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArg 142
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 AAGTTCTCTCGCGAGACACAGAGAGAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
```

```
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
Db 424 CGCGCGGCAATCAACAGCTGAGACCGCTGGCAGGACAGGTGGCGTGTATTTCTTCCT 483
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisLysGlu 182
Db 484 TCTGATGTTGGTCAGAACGCGGTAGATATCGTTAAACGGCGCTGAAAGAACGCAAACTG 543
QY 183 GluHisLeuAspPheValIleAspThrAlaGlyArgLeuHisLysLeuAspGluAlaLeu 202
Db 544 AAATCTCAGCGCTGCTGCTGGTGGTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
QY 203 MetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeuValVal 222
Db 604 ATGGACGAGATCAACAGCTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
Db 664 GACGCGCATGCGGCTCAGGATGCGGCAATACGCAAAAGCATTCATCAATGAAGCGTTAC 723
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 724 CTTACCGCGTAGTGTGACCAAGTGGACGCGGATGCCCGCGGCTGCGGCGCTCTCT 783
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
Db 784 ATTGCTCATCATCTGCAACACGATCAATCTCTCGGTGTGCGGAGAAACTGAGCGG 843
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 844 CTGGAGCGCTTCCATCCGCGCCATCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
QY 303 SerLeuIleGluLysAlaGlnLysPheValAspGlnGlnLysAlaLysAspLeuLys 322
Db 904 TGCTGATCGAAGATATATCAAAACAAAGTTGACCGCGCGCAGCAGCAAAATAGCCAGC 963
QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGlnVal 341
Db 964 AAGCTGAAAAAGGTGAGCGCTTCGATCICACAGCTTCTTGAGCAGCTGCGCCAGAG 1023
QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLys 361
Db 1024 AAAATATGGCGCATGGCTAGTCTGATGGCAAGCTGCGCGCATGGCGGCGATGCCG 1083
QY 362 GlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIleGln 381
Db 1084 GATAACGTCAGTCACAGATGGACGATAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1143
QY 382 SerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLysArg 401
Db 1144 TCGATGACGATGAAGAGCGCGCTAAGCCAGAAATCATCAAGGTTCCGCTAAACCGCT 1203
QY 402 IleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPheAsn 421
Db 1204 ATTGCTGCGGTTGCGGTATGCGAGTGCAGGTCGAGGACGTTAACCCTCTCTGAAACAGTTCGAC 1263
QY 422 AspMetLysMetMetLysGlnPheThrGlyGlyLysLysGlyLysLysArg 441
Db 1264 GACATGACGCGCATGATGAAG-----AAAATGAAGAAGGCGCGGAATG 1305
QY 442 AsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuPro 454
Db 1306 CGAAGATGATGAGACGATGAGGGTATGATGCCGCCCA 1344
```

## RESULT 13

```
US-09-815-242-1949/c
; Sequence 1949, Application US/09815242
; Patent No. US20020061589A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1949
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-1949
```

## Alignment Scores:

Pred. No.:	2,4e-121	Length:	721
Score:	1158.00	Matches:	234
Percent Similarity:	99.58%	Conservative:	2
Best Local Similarity:	98.73%	Mismatches:	1
Query Match:	50.92%	Indels:	0
DB:	10	Gaps:	0

US-09-943-108a-2 (1-455) x US-09-815-242-1949 (1-721)

```
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 713 ATGCGATTTGAGGGTATCAGACGCTTCAAGCGACGATGCAAAAAATCGCTGGTAAG 654
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 653 GGTAAACTTTACTGAAAGCTGATATAAGATAATGATCGCTGAAGTAAGATTAGCGTTACTT 594
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 593 GAGCGTACGTAAACTTTAAAGTGGTAAAGAAATTTATTAAACAGTATCAGAACCGCA 534
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 533 TTAGTTCCGATGTAATCAATCATTAACACACGAGCAACAAAGTTATTAAATAGTTCAA 474
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 473 GATGAATTAACACAGTTCGTTGGTGGAGAAATACGCGATTATATGTCAAATAACCA 414
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 413 CCTACTGTTGTTATGATGGTTGGTTTACAAAGGTCGTGGTAAACAAACAACTGCAGGTAA 354
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 353 TTAGCATTTATGTCGTTAAANAATACACAAAAACCTNTGTTAGTTGCAGCAGATATT 294
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 293 TATCGTCGACGACGATTAATCAATTAACAAACAGTAGGGAACAAATGTATATTCCTGTA 234
```

161	TyrSerGluCluYasGlnValLysProGInGlnIleValThrAsnAlaLeuLysHisAla	180
QY		
233	TACAGTGAAGGAGATCAAGTAAGACGCACACAAATGTGAACATAATGCATTAATAACATGCT	174
Db		
181	LysGluCluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu	200
QY		
173	AAAGAGACACATTAGACTTTGTAATCATTGATACAGAGGTGATACACATCGATGAA	114
Db		
201	AlaLeuMetAsnGluLeuLysGluValLysGlnIleAlaLysProAsnGluIleMetLeu	220
QY		
113	GCATTGATGACGAATTAAGAGAGTAAGAAGACATTCCTAAACCAACGAANTATATGTTA	54
Db		
221	ValValAspSerMetThrGlyGlnAspAlaValAsnValIlaGluSerPhe	237
QY		
53	GTTCGATTCATTAACGCGGTCAAGATGCTGCAATGTTCGAAATCTTTT	3
Db		

RESULT 14

US-09-815-242-7850  
; Sequence 7850, Application US/09815242

; FACILE NO: US2002006;  
; GENERAL INFORMATION:  
; GENERAL INFORMATION:

```

/ APPLICANT: Haseibeck, Robert
/
/ APPLICANT: Ohlsen, Karl L.
/
/ APPLICANT: Zyskind, Judith W.
/
/ APPLICANT: Wall, Daniel
/
/ APPLICANT: Trawick, John D.
/
/ APPLICANT: Carr, Grant J.
/
/ APPLICANT: Yamamoto, Robert T.
/
/ APPLICANT: Xu, H. Howard
/
/ TITLE OF INVENTION: Identification of Essential Genes in
/
/ TITLE OF INVENTION: Prokaryotes
/

```

```

/ CURRENT APPLICATION NUMBER: US/09/815,242
/
/ CURRENT FILING DATE: 2001-03-21
/
/ PRIOR APPLICATION NUMBER: 60/191,078
/
/ PRIOR FILING DATE: 2000-03-21
/
/ PRIOR APPLICATION NUMBER: 60/206,848
/
/ PRIOR FILING DATE: 2000-05-23
/
/ PRIOR APPLICATION NUMBER: 60/207,727
/
/ PRIOR FILING DATE: 2000-05-26
/
/ PRIOR APPLICATION NUMBER: 60/242,578
/
/ PRIOR FILING DATE: 2000-10-23
/
/ PRIOR APPLICATION NUMBER: 60/253,625
/
/ PRIOR FILING DATE: 2000-11-27
/
/ PRIOR APPLICATION NUMBER: 60/257,931
/
/ PRIOR FILING DATE: 2000-12-22
/
/ PRIOR APPLICATION NUMBER: 60/269,308
/

```

```

/ ** BUILD INFORMATION **/
/ ** NUMBER OF SEQ ID NOS: 14110 **/
/ ** SOFTWARE: FastSeq for Windows Version 4.0 **/
/ ** SEQ ID NO 7850 **/
/ ** LENGTH: 1374 **/
/ ** TYPE: DNA **/
/ ** ORGANISM: Pseudomonas aeruginosa **/
/ ** FEATURE: **/
/ ** NAME/KEY: CDS **/
/ ** LOCATION: (1)...(1374) **/
US-09-815-242-7850

```

Alignment Scores:

Argument Scores:		
Pred. No.:	4.79e-120	Length: 1374
Score:	1150.00	Matches: 231
Percent Similarity:	70.90%	Conservative: 93
Best Local Similarity:	50.55%	Mismatches: 121
Query Match:	50.57%	Indels: 12
DB:	10	Gaps: 3

US-09-943-108A-2 (1-455) X US-09-815-242-7850 (1-1374)

QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22  
||||| |||||:::||||| |||||::: |||||::: ||||| |||||

Db 4 TTTCGAAACCTTACAGACCGCCCTCTCGCAGACGCTTCGCATGTCACCGCAAGSCCAAG 63

QY	23	LeuthrGluAlaSpIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla	42
DB	64	CTGACCGAGGACAACATCAAGAGACACTCTGGCGGAGTGGCATGGCCCTGCTCGAGGCC	123
QY	43	AspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAlaLeuGly	62
DB	124	GACGTGGCCCTCCGGTGGTCAAGGACTTCGTCAACAGGTCAAGGACGCGCGTCCGGT	183
QY	63	SerAspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGlnAspGlu	82
DB	184	ACCGAGGTCTCAGAGACCTTGACCCGGGACGAGCGTTCGTGAAGATGTCGCGCGCCGAG	243
QY	83	LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProProThr	102
DB	244	CTCGAGGAGTGATGGGGCGGCCAACAGGACCTGCGCGTGGCGTGGCGCGCGCGCG	303
QY	103	ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla	122
DB	304	GTGATCTGATGCGCGCCCTGCAGGCGCGGGCAGACACACCGCGGGCAGAGTGGCG	363
QY	123	LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArg	142
DB	364	CGCTTCCTTAAGAACCGCAGAGACAGTGGTGATGGTGGTTCGCGCGAGCTTCAACGA	423
QY	143	ProAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer	162
DB	424	CCGCGCGCTATCAAGCATGGAAACCCCTGGCGGCGAGTGGCGGTAACTTCTTCCT	483
QY	163	GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu	182
DB	484	TCCGACGTCACGACGAAGCGGGTGGCAATCGCGGCGAGCGGCATCCGCGAGGCGAGG	543
QY	183	GluHisLeuAsnPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu	202
DB	544	ASGTTCAATCGACGTGGTGTGTCATCCGAGCGCGCTGCATCATCATGCGCCAGTGC	603
QY	203	MetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValVal	222
DB	604	ATGACGAGATCAAGCAGGTGCACGCGGCGATCAACGCGCGGAAACCCCTGTCGTGTC	663
QY	223	AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp	242
DB	664	GATGCCATGACCGCGCGAGATGCGGCCAACCGCCAGGCGCTTCATGACGCGCTGGCG	723
QY	243	ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer	262
DB	724	CTGACCGCGGTGCTCTCACCAAGGTCGACGCGCGCGTGGTGGTGGCGCGCTCTCG	783
QY	263	IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly	282
DB	784	GTGGCGGCGATCACCGGACGCGATCAAGTTCCTCGGCATCGGCGCAAGACGCGAAGCG	843
QY	283	LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu	302
DB	844	CTCGATCCGTTCATCCGCGCGGTGGCTCGCGATCTCTGGCATGGGCGAGCTGCTC	903
QY	303	SerLeuIleGlyAlaGlnGlnAspValAspGlnGlyLysAlaLysAspLeuLys	322
DB	904	AGCTGATCGACAGCGCGAGAACCTTCGACCGCAGACAGCGCGAAGCTGGCGAAG	963
QY	323	LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnVal	341
DB	964	AAGATCAAGAGGCAAGGGCTTCGACCTGGAAACATCTCGGAGCACCATGATGCAACAGATG	1023
QY	342	LysAsnLeuGlyProLeuAspAspIleMetLysMetIlePro-----GlyMetAsn	358
DB	1024	AAGAACATGGCGCGCTCGCGCGCTCATGACAAAGTCCCATGCTCGCGGGGCTCAAC	1083
QY	359	LysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAla	378
DB	1084	CTGGCGCAGTGGCCAAATGATGCCAGGTGCTCGGAGAAACAGTCAACGACAGATGGAGCG	1143
QY	379	IleIleGlnSerMetThrProAlaGluArgAsnAsnProAsnThrLeuAsnValSerArg	398



```
Db 1144 ATCATCACTCGATGACCGCCGGCGAAGCGGATCGGGAATCATCAGCGCTCGGC 1203
QY 399 LysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAlaSerLeuMetLys 418
Db 1204 AAGCGCGCATCGCCCTTGGCTCCGGTACCCAGGTGCAGGCGTGGCGGCTGATCAG 1263
QY 419 GlnPheAsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLys 438
Db 1264 CAGCACAGCAGATGACAGAAATGATGAGAAGGTCACCGCAAGGCG----- 1311
QY 439 GlyLysArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1312 -----GGCATGGCCCAAGATGATCGGAGGCGTGGCGATGATGTTTC 1350

RESULT 15
US-09-815-242-6868
; Sequence 6868, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlisen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6868
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1389)
US-09-815-242-6868

Alignment Scores:
Pred. No.: 1,09e-118 Length: 1389
Score: 1138.00 Matches: 225
Percent Similarity: 70.67% Conservative: 93
Best Local Similarity: 50.00% Mismatches: 122
Query Match: 50.04% Indels: 10
DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-815-242-6868 (1-1389)
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTTGAGATTATTCGATCGCGCTTCCAAACCTTACGTATATACAGGAAGAGCGCT 63
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
```

```
Db 64 TTACCGAAGATAATATTAAGAAACCTTACCGAAGTGGCTATGCACTTACTTGAAGCC 123
QY 43 AspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db 124 GATGTTGCCCTTGGCTGGTGGTGAATTTATCGCAAAAGTAAAGAAAGCGCGTGGG 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysValGlnAspGlu 82
Db 184 GAAGAAGTCAATAAAGTTTAAGCCAGGCAAGAAATCTTAAAAATCGTTCAGCGTGAG 243
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
Db 244 CTTGAAAAAGCCATGGGCGAAGCAATGAGCTTTAAATCTCGCAACCCACCAACGACGA 303
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTTATCTTAATGGCGGTTTACAAGGGCGGCTAAACACCACGATGGTGAATTTGGCA 363
QY 123 LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleLysArg 142
Db 364 AAATCTTCCGTGAACGCCCATAAAAAGAAAGTTAGTGGTGTCTGCTGACGTATATCGC 423
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
Db 424 CTTGCTGGCATTAAGCACTTGAACCTTGGCTCAATCCGTGGCGGATTTTTCCTCA 483
QY 163 GluGlyAspGlnLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
Db 484 TCGGATGTTAAACAAAAACCCGCTTGATATTGCTAAATCGGCGCTGCTGATGCAAACTG 543
QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu 202
Db 544 AAATCTACGATGTTGATGTTGGTACGCGAGTCCCTTACACCTTATACAGAAATG 603
QY 203 MetAsnGlnLysLysGluValLysGluIleAlaLysProAsnGlnIleLeuValVal 222
Db 604 ATGGAGAAATCAAGCAAGTCCATGCTGCATTAATCAATCGAAACTCTTTTCACTGT 663
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
Db 664 GATGCGTACGTGCTCAAGATGGCGCAATACAGCAAAAGCTTTAATGAAGCATTGCT 723
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 724 CTTACAGGCGTTATTTGACAAAGTGGACGGTATGCGCGCGGTGGTGGCGGTATTCG 783
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeuAspGly 282
Db 784 ATTCGCAAAATCACAGGTAAACCAATCAAAATCTTGGGTGPGGCGAGAAAAACAGAGCG 843
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 844 CTTGAGCCATTCCATCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnLysAlaLysAspLeuLys 322
Db 904 TCCCTTATCGAAGATCTTGAACGTTCTGTTGATCTGAAAGAGCGGAAAAAATGGCGGAG 963
QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGlnLeuAspGlnVal 341
Db 964 AAATTCAGAAAGCGCATGATGATTTTACTTTAGATGATTTCCGCAACAGCTGATGAAATG 1023
QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLys 361
Db 1024 AAAAAATGGCGCGCATGATGCTATGCTGAGAAATATCCAGGTGCAAAAAAATTTGCT 1083
QY 362 GlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLeuGln 381
Db 1084 GACACGTTAAATAATCAAGTAGATGACAAAAAGTTTGTCAAAATGGAGCGATCATTAAC 1143
QY 382 SerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLysArg 401
```

```

Db 1144 TCCATGACCCCTAAAGAACGTGCCAACCCAGATATTATCAAGAGATCTCGCCGTCTCGT 1203
Qy 402 IleAlaLysGlySerClyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPheAsn 421
Db 1204 ATGGCATTAGGCTCTGGCACTCAAGTCGAAGATGTCATTAATTACTTAACAACTTCGAT 1263
Qy 422 AspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLysArg 441
Db 1264 GAATGCACGTATGATGAAGAAATGCGTAAGGCGGC----- 1302
Qy 442 AsnGlnMetGlnAsnMetLeuLysGlyMet 451
Db 1303 -----ATGGCTAAAATGATGCGTGAATG 1326

```

Search completed: February 25, 2003, 04:03:39  
Job time : 113 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: February 21, 2003, 15:16:32 : Search time 22 Seconds  
(without alignments)  
1988.235 Million cell updates/sec

Title: US-09-943-108a-2  
Perfect score: 2274  
Sequence: 1 MAFEGLSRLQATMQRGK.....GKKGRNOMNMLKGMNLPF 455

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2274	100.0	455	2 H89896	hypothetical prote
2	1717.5	75.5	450	2 A11299	signal recognition
3	1708.5	75.1	450	2 A11671	signal recognition
4	1647	72.4	446	2 B47154	signal recognition
5	1584	69.7	451	2 D83960	signal recognition
6	1393	61.3	449	2 E97116	signal recognition
7	1315	57.8	518	2 H86826	signal recognition
8	1307	57.5	523	2 E98017	hypothetical prote
9	1306	57.4	523	2 F95149	signal recognition
10	1180.5	51.9	433	2 F72236	signal recognition
11	1173.5	51.6	453	2 A10833	signal recognition
12	1164.5	51.2	453	2 E65039	signal recognition
13	1164.5	51.2	453	2 A91063	signal recognition
14	1164.5	51.2	453	2 E95907	signal recognition
15	1150.5	50.6	461	2 H83306	signal recognition
16	1150	50.6	457	2 F83178	signal recognition
17	1139.5	50.1	453	2 A04000	signal recognition
18	1138	50.0	462	2 H64048	signal recognition
19	1131	49.7	456	2 F81215	signal recognition
20	1130	49.7	456	2 B1792	signal recognition
21	1122.5	49.4	482	2 S75847	signal recognition
22	1106.5	48.7	496	2 S68160	probable RNA bindi
23	1092.5	48.0	490	2 A82050	signal recognition
24	1091.5	48.0	525	2 D70747	probable fth prote
25	1077.5	47.4	550	2 H37711	signal recognition
26	1069	47.0	521	2 H87111	signal recognition
27	1066.5	46.9	452	2 G75346	signal recognition
28	1062.5	46.7	523	2 H2907	signal recognition
29	1062.5	46.7	554	2 H97682	signal recognition

ALIGNMENTS

RESULT 1

H89896  
hypothetical protein fth [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: H89896  
R:kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.;  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu,  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89756; MUID:21311952; PMID:11418146  
A:Accession: H89896  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <R>  
A:Cross-references: GB:BA000018; PID:g13701037; PIDN:BAB42332.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: fth  
C:Superfamily: signal recognition particle 54K protein

Query Match	100.0%	Score 2274;	DB 2;	Length 455;
Best Local Similarity	100.0%	Pred. No. 7.6e-109;	Mismatches 0;	Gaps 0;
Matches 455;	Conservative 0;			
QY 1	MAFEGLSRLQATMQRGKGLTEADIKIMREVLALFEADVNFVKVKEFIKTVSERA 60			
DB 1	MAFEGLSRLQATMQRGKGLTEADIKIMREVLALFEADVNFVKVKEFIKTVSERA 60			
QY 61	LGSDVMGSLTPGOVTKIVQDELTKMGGENTSNMKNKPPVTVMVVGLOGAGKTTAGK 120			
DB 61	LGSDVMGSLTPGOVTKIVQDELTKMGGENTSNMKNKPPVTVMVVGLOGAGKTTAGK 120			
QY 121	LALLMRKYNKPMVAADYRPAAINQLQTVGKQIDIPVYSEGQVKKPQIVTNALKEA 180			
DB 121	LALLMRKYNKPMVAADYRPAAINQLQTVGKQIDIPVYSEGQVKKPQIVTNALKEA 180			
QY 181	KEEHLDFVIIDTAGRLHDEALMNELKEVKEIAKPMELVYDSMTGDAVNVAFESDDQ 240			
DB 181	KEEHLDFVIIDTAGRLHDEALMNELKEVKEIAKPMELVYDSMTGDAVNVAFESDDQ 240			
QY 241	LDVTVGVTFLKDGDFGGAALSIRSVTKPKIFVGMSEKLDGLLELFPERWASRIILGMD 300			
DB 241	LDVTVGVTFLKDGDFGGAALSIRSVTKPKIFVGMSEKLDGLLELFPERWASRIILGMD 300			
QY 301	VLSLIEKAQDVQDEKAKDLKKMRESSTLDDFLQDLQVKNLGLPDDIMKIPGNMKN 360			
DB 301	VLSLIEKAQDVQDEKAKDLKKMRESSTLDDFLQDLQVKNLGLPDDIMKIPGNMKN 360			
QY 361	KGLDKLNMSEKQIDHIIKAILOSMTPAERNPDPLVNSRKKRIAGSGRSLOEVNRLMKQF 420			



.....

A;Reference number: A83850; MUID:20512582; PMID:11058132  
A;Accession: D83960

100

Db 1 MAFEGSLTQLOAMKGLRGLKSLSEKIDKAMREYKLLALLADYVNIYKFNFKVVGK 60  
QY 61 LGSVMQSLTPGOQVIVQDELTKLGGGENTSIINMSKPPYVVMVGLQGAGKTTTACK 120  
Db 61 LGNEVMESLTPGOQVIVQDELTKLGGGENTSIINMSKPPYVVMVGLQGAGKTTTACK 120  
QY 121 LALLMEKKYKPMVAADYRPAALNOLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
Db 121 LLSLSLKK-NKKPLLVACDIYRPAALNOLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 179  
QY 181 KEEHDFVIIDTAGRLHIDEALMNEKVEKTAAPNEIMLVVDSMTGQDAVNVAFSDQ 240  
Db 180 KENGLNVVIDTAGRLHIDEALMNEKVEKTAAPNEIMLVVDSMTGQDAVNVAFSDQ 239  
QY 241 LDVTGVTLLKLDGTRGGAALSRVSTQKPIKFGVMSKLDGLFELFPERMASRILGMGD 300  
Db 240 LETGVVLTFLKLDGTRGGAALSRVSTQKPIKFGVMSKLDGLFELFPERMASRILGMGD 299  
QY 301 VLSLIEKAQDVDEKAKDLKKMRESSTLDDLEQLDQVKNLGPLDDIMKMPGNK- 359  
Db 300 VLILIEKAQSAIDEKQAKELGDRMLSQEFNFDLQAEQMKLGPGLKLEWVFGFNSS 359  
QY 360 -MKGLDKLMSKQIDHIKALIOSMTPAERNPD--TLNVRKKRIAKGSRSLQEVNRL 416  
Db 360 MLKAGVD-LSKNGEKKYKPMVAADYRPAALNOLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 418  
QY 417 MKQFDMKKMKQFTGGGKGGKGRKQNMQLKGMNLP 455  
Db 419 LKNEQMKMKQFTGGGKGGKGRKQNMQLKGMNLP 449

RESULT 7  
H86826  
signal recognition particle protein Ffh [imported] - Lactococcus lactis subsp. lactis (S  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: H86826  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A:Reference number: A86625; MUID:21235166; PMID:11337471  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-518 <STO>  
A:Cross-references: GB:AE005176; PID:g12724623; PIDN:RAK05714.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: ffh  
C:Superfamily: signal recognition particle 54k protein

Query Match 57.8%; Score 1315; DB 2; Length 518;  
Best Local Similarity 55.4%; Pred. No. 6.1e-60;  
Matches 258; Conservative 88; Mismatches 102; Indels 18; Gaps 2;

QY 1 MAFEGSLTQLOAMKGLRGLKSLSEKIDKAMREYKLLALLADYVNIYKFNFKVVGK 60  
Db 1 MAFENLTERLQNVFKNLKRGKKITETDVTETKIRVALLEADVALPVVKFKTKAIRERA 60  
QY 61 LGSVMQSLTPGOQVIVQDELTKLGGGENTSIINMSKPPYVVMVGLQGAGKTTTACK 120  
Db 61 VGVEVSALNPAAQVQVIVQDELTKLGGGENTSIINMSKPPYVVMVGLQGAGKTTTACK 120  
QY 121 LALLMEKKYKPMVAADYRPAALNOLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
Db 121 LAKLKEBQNAFLMAADYRPAALNOLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
QY 181 KEEHDFVIIDTAGRLHIDEALMNEKVEKTAAPNEIMLVVDSMTGQDAVNVAFSDQ 240  
Db 181 QEEKDYVLIDTAGRLHIDEALMNEKVEKTAAPNEIMLVVDSMTGQDAVNVAFSDQ 240  
QY 241 LDVTGVTLLKLDGTRGGAALSRVSTQKPIKFGVMSKLDGLFELFPERMASRILGMGD 300  
Db 241 LDVTGVTLLKLDGTRGGAALSRVSTQKPIKFGVMSKLDGLFELFPERMASRILGMGD 300

Db 241 LDVTGVTLLKLDGTRGGAALSRVSTQKPIKFGVMSKLDGLFELFPERMASRILGMGD 300  
QY 301 VLSLIEKAQDVDEKAKDLKKMRESSTLDDLEQLDQVKNLGPLDDIMKMPGNK 360  
Db 301 MLTLIEKAQANYDEQSAKLAEKANRFDYEDFVEQLDQVNNNGPMDIMKMPGNM 360  
QY 361 KGLDKLMSKQIDHIKALIOSMTPAERNPD--TLNVRKKRIAKGSRSLQEVNRLKQF 420  
Db 361 PGLDKVYKDPKDVARKKAWYLVSTPAERHLEAEUSPARRRIAGSGNSFIEVNFKIQF 420  
QY 421 N-----DMKKMKQFTGGGKGGKGRKQNMQLKGMNLP 454  
Db 421 NOSKEMAGIANGDNAMQKMGAGG-----QMPNMPAGSGMP 460

RESULT 8  
E98017  
hypothetical protein ffh [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 17-May-2002  
C:Accession: E98017  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.;  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren,  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, I  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: E98017  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-523 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99969.1; PID:g15458797; GSPDB:GN00174  
C:Genetics:  
A:Gene: ffh  
C:Superfamily: signal recognition particle 54k protein

Query Match 57.5%; Score 1307; DB 2; Length 523;  
Best Local Similarity 55.1%; Pred. No. 1.6e-59;  
Matches 253; Conservative 90; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGSLTQLOAMKGLRGLKSLSEKIDKAMREYKLLALLADYVNIYKFNFKVVGK 60  
Db 1 MAFESLTERLQNVFKNLKRGKKITETDVTETKIRVALLEADVALPVVKFKTKAIRERA 60  
QY 61 LGSVMQSLTPGOQVIVQDELTKLGGGENTSIINMSKPPYVVMVGLQGAGKTTTACK 120  
Db 61 VGHEVIDTLNPAQQLIKIVDELTAVIGSDTAETIKSPKPTIIMVGLQGAGKTTTACK 120  
QY 121 LALLMEKKYKPMVAADYRPAALNOLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
Db 121 LANKLKEBQNAFLMAADYRPAALNOLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
QY 181 KEEHDFVIIDTAGRLHIDEALMNEKVEKTAAPNEIMLVVDSMTGQDAVNVAFSDQ 240  
Db 181 QTHNDYVLIDTAGRLHIDEALMNEKVEKTAAPNEIMLVVDSMTGQDAVNVAFSDQ 240  
QY 241 LDVTGVTLLKLDGTRGGAALSRVSTQKPIKFGVMSKLDGLFELFPERMASRILGMGD 300  
Db 241 LEVTGVTLLKLDGTRGGAALSRVSTQKPIKFGVMSKLDGLFELFPERMASRILGMGD 300  
QY 301 VLSLIEKAQDVDEKAKDLKKMRESSTLDDLEQLDQVKNLGPLDDIMKMPGNK 360  
Db 301 MLTLIEKASQYDEQSAKLAEKANRFDYEDFVEQLDQVNNNGPMDIMKMPGNM 360  
QY 361 KGLDKLMSKQIDHIKALIOSMTPAERNPD--TLNVRKKRIAKGSRSLQEVNRLKQF 420  
Db 361 PALQNMKVDHETAKKALVSSMTPEERNPOLLNPSRRRIAGSGNSFIEVNFKIQF 420  
QY 421 NDMKKMKQFTGGGKGGKGRKQNMQLKGM-----NLP 454  
Db 421 NQAKQLAQGVMSG-----DMKKMKQMGINPNLP 450

A;Accession: F72236  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-433 <ARN>  
A;Cross-references: GB:AE001802; GB:AE000512; NID:94982133; PIDN:AAD36632.1; PID:q4.94  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TMJ565  
C;Superfamily: signal recognition particle 54K protein

Query Match 51.9%; Score 1180.5; DB 2; Length 433;  
Best Local Similarity 53.4%; Pred.No. 3.4e-53;  
Matches 234; Conservative 92; Mismatches 99; Indels 13; Gaps 5;

QY 3 PEGLSERLOATMKMGCKGLTEADIKLMREVRLLAFADVNEKVVKFEIKTVSERALG 62  
||| : : |||:|||| : : |||:|||| : |||:|||| : |||:|||| : |||:||||  
DB 2 FENLQELSRVFNLSGRGKITERNKDAIREVKSLJEADVNTKKVKEFVDHVIQKALG 61  
:||||| || |||||:||| : : : : : |||||:||||| |||  
QY 63 SDVMQSITPQQQVIKTIVDELTKLGGENTISNMSPPTVVMMVGVLQGAGKTTTAKLIA 122  
DB 62 BEVLSUTPDQQFIKIVRDELVRIMGKEPNELRVHR-PAPIMAVGIQSGSKTTTCAMIA 120  
||| : : |||:||||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
QY 123 LLMRRKNKXPMVAADIYRPAAINQLQTGVKGIDIPVVSQGDQVKKQIVTWALKHAKE 182  
||| : : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
DB 121 KLL-KKEGRNELVAADLYRPAADVQLVGLGNIGVNVVHDYINKT-EVEIVKEADVDAES 178  
:||||| || |||||:||| : : : : : |||||:||||| |||  
QY 183 EHLDFVIDTAGRLHIDEALMNEIKVEKTAQNEMLVDSMTGQDANVAESFDQILD 242  
DB 179 TGSQDVLIIVDTAGRLHIDEMKLEELTKILNPDEILLVDAKAGQDAVNTKVPDERLD 238  
||| : : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
QY 243 VTGVTLFKLOGTRFGAALSTSVTPQPIKFVGNSEKLDGLELHPHERMASRIILGMGDLV 302  
||| : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
DB 239 LTGFVTWKMDQARGVILSIKYVTGPKVFIFGTSEKLDGLEPFPHDPRIANRIILGMGDLV 298  
||| : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
QY 303 SLIEKAQDDVDQSKADLEKKMRSSSTLDDFTLEQLDDQVKNLGPLDIMKMPGNMKMG 362  
DB 299 SLIEKVENLDQEKMKKSAREKFLRAEFTLDFEQLEQMKGKLGSLILLEMLPGAPKV-- 356  
||| : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
QY 363 LDKLNASEKIDIKAIIOGMTPPARNNPDTLVNSKRRIARKSGRSLOEVNLRMKQFND 422  
||| : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
DB 357 --DVMESEKELKXIEALINSNTIERRNPGIIINASRKRIARGSGTTVDVYNKLKASYEQ 414  
||| : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
QY 423 MKKMKMQFTGGKGGKKGK 440  
||| : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||  
DB 415 MKALMKRM-----KXGR 426  
||| : |||:|||||:||| : |||:||||| : |||:||||| : |||:||||| : |||:||||

RESULT 11  
AI0833  
signal recognition particle protein [imported] - Salmonella enterica subsp. enterica  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: This species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AI0833  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church,  
T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Teague,  
S.; Moulder, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parkhill, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A.; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica  
A;Reference number: AB0502; PMID:11677608  
A;Accession: AI0833  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-453 <FAR>  
A;Cross-references: GB:AL513382; PIDN:CAD05856.1; PID:gl6503831; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY2864  
C;Superfamily: signal recognition particle 54K protein

Query Match 51.6%; Score 1173.5; DB 2; Length 453;  
Best Local Similarity 51.4%; Pred.No. 8.2e-53;  
Matches 233; Conservative 93; Mismatches 120; Indels 7; Gaps 2;

```
QY 3 FEGLSERLOATMOKMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTSERALG 62
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 2 FDNLTDLRLSRLNISRGRLETDNVKTLREVRMALLEADVALPVVREFINRVKEKAVG 61
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 63 SDVMOSLTPGQGVKIVQDELTKLMGGENTSINMSKPPVVMVMGLOGAGKTTTAGKLA 122
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 62 HEVNSLTPGQGVKIVRSELVAAMGEENQTLNLAAQPPAVVLMAGLOGAGKTTSVGKLG 121
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 123 LLMRKKYKPKMLVAADIVRPAAINQLOTVGKQIDIPVYSEGQVVKPOQIVTNALKHAKE 182
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 122 KFLREKHKKKVYVVSADVYRPAAIKQLETLAEQVGVDFPDSVGVKQKPDVIVNAALKEAKL 181
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 183 EHLDFVIIDTAGRLHIDEALMNELKEVEIAKPNEIMLVVDSMTGODAVNVAESFDDQLD 242
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 182 KEYDVLVDTAGRLHIDEAMDEIKQVHASINPVETLFFVDMTGDAAANTAKAFNEALP 241
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 243 VTGVTLTCLDGTDRGGAALSIRSVTKPKIFVGMSEKLDGLELFFHPERMASRIILGMDVL 302
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 242 LTGVVLTAKVDGDARGGAALSIRHITGPKIFLGVEKTEALEPFPHPDRIASRIILGMDVL 301
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 303 SLIEKAQDVDOEKAKDLEKKMRE-SSFTLDDLEQLDQVKNLGPLDDIMKMIKPMKMK 361
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 302 SLIEDIESKVDRAQAELKATLKKGDGFDLNDLEQLQKMGNGGWSLWGLKLPFGMGQIP 361
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 362 GLDKLNKSEKQIDHIIKAIQSMTPAERNPDTLVNSRKKRIAGSGRSLOEVNRLMKQFN 421
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 362 DNKSKQMDKVLVPMERAINLSMTKERAKPEIIGKSRKRIAGCGMQVQDVNRLLKQFD 421
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 422 DNKSKMKGFTGGGKGGKGRNQMNKLGMINLP 454
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 422 DMQRMMK-----KMKKGGKAKMRSKMGMP 448
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |

RESULT 12
E65039
signal recognition particle protein - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: E65039; S07178
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E65039
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-453 <BLAT>
A:Cross-references: GB:AE000347; GB:U00096; NID:g2367142; PIDN:AAC75659.1; PID:g1788963;
A:Experimental source: strain K-12, substrain MG1655
R:Bystroem, A.S.; Hjalmarsson, K.J.; Wikstroem, P.M.; Bjoerk, G.R.
EMBO J. 2, 899-905, 1983
A:Title: The nucleotide sequence of an Escherichia coli operon containing genes for the
A:Reference number: A30380; MUID:84057772; PMID:6357787
A:Accession: S07178
A:Molecule type: DNA
A:Residues: 86-453 <BYS>
A:Cross-references: EMBL:X01818; NID:g43141; PIDN:CAA25957.1; PID:g43142
C:Genetics:
A:Gene: ffh
A:Map position: 57 min
C:Superfamily: signal recognition particle 54k protein

Query Match 51.2%; Score 1164.5; DB 2; Length 453;
Best Local Similarity 51.0%; Pred. No. 2.4e-52;
Matches 231; Conservative 94; Mismatches 121; Indels 7; Gaps 2;

QY 3 FEGLSERLOATMOKMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTSERALG 62
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 2 FDNLTDLRLSRLNISRGRLETDNVKTLREVRMALLEADVALPVVREFINRVKEKAVG 61
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 63 SDVMOSLTPGQGVKIVQDELTKLMGGENTSINMSKPPVVMVMGLOGAGKTTTAGKLA 122
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 122 KFLREKHKKKVYVVSADVYRPAAIKQLETLAEQVGVDFPDSVGVKQKPDVIVNAALKEAKL 181
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 183 EHLDFVIIDTAGRLHIDEALMNELKEVEIAKPNEIMLVVDSMTGODAVNVAESFDDQLD 242
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 182 KEYDVLVDTAGRLHIDEAMDEIKQVHASINPVETLFFVDMTGDAAANTAKAFNEALP 241
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 243 VTGVTLTCLDGTDRGGAALSIRSVTKPKIFVGMSEKLDGLELFFHPERMASRIILGMDVL 302
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 242 LTGVVLTAKVDGDARGGAALSIRHITGPKIFLGVEKTEALEPFPHPDRIASRIILGMDVL 301
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 303 SLIEKAQDVDOEKAKDLEKKMRE-SSFTLDDLEQLDQVKNLGPLDDIMKMIKPMKMK 361
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 302 SLIEDIESKVDRAQAELKATLKKGDGFDLNDLEQLQKMGNGGWSLWGLKLPFGMGQIP 361
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 362 GLDKLNKSEKQIDHIIKAIQSMTPAERNPDTLVNSRKKRIAGSGRSLOEVNRLMKQFN 421
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 362 DNKSKQMDKVLVPMERAINLSMTKERAKPEIIGKSRKRIAGCGMQVQDVNRLLKQFD 421
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 422 DNKSKMKGFTGGGKGGKGRNQMNKLGMINLP 454
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 422 DMQRMMK-----KMKKGGKAKMRSKMGMP 448
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |

RESULT 12
A91063
signal recognition particle protein Ffh [imported] - Escherichia coli (strain O157:H
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: A91063
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han,
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: A91063
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-453 <GB>
A:Cross-references: GB:BA000007; PIDN:BA36896.1; PID:g13362944; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs3473
C:Superfamily: signal recognition particle 54k protein

Query Match 51.2%; Score 1164.5; DB 2; Length 453;
Best Local Similarity 51.0%; Pred. No. 2.4e-52;
Matches 231; Conservative 94; Mismatches 121; Indels 7; Gaps 2;

QY 3 FEGLSERLOATMOKMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTSERALG 62
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 2 FDNLTDLRLSRLNISRGRLETDNVKTLREVRMALLEADVALPVVREFINRVKEKAVG 61
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 63 SDVMOSLTPGQGVKIVQDELTKLMGGENTSINMSKPPVVMVMGLOGAGKTTTAGKLA 122
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 62 HEVNSLTPGQGVKIVRSELVAAMGEENQTLNLAAQPPAVVLMAGLOGAGKTTSVGKLG 121
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 123 LLMRKKYKPKMLVAADIVRPAAINQLOTVGKQIDIPVYSEGQVVKPOQIVTNALKHAKE 182
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 122 KFLREKHKKKVYVVSADVYRPAAIKQLETLAEQVGVDFPDSVGVKQKPDVIVNAALKEAKL 181
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 183 EHLDFVIIDTAGRLHIDEALMNELKEVEIAKPNEIMLVVDSMTGODAVNVAESFDDQLD 242
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 182 KEYDVLVDTAGRLHIDEAMDEIKQVHASINPVETLFFVDMTGDAAANTAKAFNEALP 241
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
QY 243 VTGVTLTCLDGTDRGGAALSIRSVTKPKIFVGMSEKLDGLELFFHPERMASRIILGMDVL 302
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 242 LTGVVLTAKVDGDARGGAALSIRHITGPKIFLGVEKTEALEPFPHPDRIASRIILGMDVL 301
|: : : | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
```



Search completed: February 21, 2003, 15:19:26  
Job time : 24 secs

RESULT 15  
H82306

GenCore version 5.1.3

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:13:26 ; Search time 14 Seconds

(without alignments)  
1347.981 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274

Sequence: 1 MAFEGLSERLQATMKMRGK.....GKKGRNQNMVLKGNLFF 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	72.4	446	1 SR54_BACSU	P37105 bacillus su
2	1286	56.6	516	1 SR54_STRMU	Q54431 streptococc
3	1164.5	51.2	453	1 SR54_ECOLI	P07019 escherichia
4	1138	50.0	462	1 SR54_HARIN	P44518 haemophilus
5	1122.5	49.4	482	1 SR54_SYNY3	P74214 synchocyst
6	1106.5	48.7	485	1 SR54_SYNP7	Q55311 synchococc
7	1091.5	48.0	525	1 SR54_MYCTU	Q33013 mycobacteri
8	1069	47.0	521	1 SR54_MYCLE	P37107 arabidopsis
9	1023	45.0	564	1 SR5C_ARATH	O67615 aquifex aso
10	1009	44.4	454	1 SR54_AQUAE	P57473 bacterera ap
11	1006	44.2	451	1 SR54_BUCAI	O07347 thermus aqu
12	1005.5	44.2	429	1 SR54_THBAQ	Q9zd20 rickettsia
13	958.5	42.2	449	1 SR54_RICPR	Q01442 mycoplasma
14	944	41.5	447	1 SR54_MYCMY	P47294 mycoplasma
15	898	39.5	446	1 SR54_MYCGE	P75054 mycoplasma
16	895	39.4	450	1 SR54_MVCPN	Q9zk62 helicobacte
17	854	37.6	448	1 SR54_HELPJ	P56005 methanococc
18	848	37.3	448	1 SR54_HELPJ	Q57565 methanococc
19	787	34.6	451	1 SR54_METJA	O07853 sulfolobus
20	761.5	33.5	446	1 SR54_SULAC	Q97ze7 pyrococcus
21	759.5	33.4	443	1 SR54_PYRAB	O59307 pyrococcus
22	748.5	32.9	443	1 SR54_PYRHO	P70722 acidiana a
23	723.5	31.8	451	1 SR54_ACTAM	O27376 methanobact
24	721	31.7	442	1 SR54_METHH	Q99633 archaeoglob
25	701.5	30.8	433	1 SR54_ARCFU	Q97ze7 sulfolobus
26	695.5	30.6	447	1 SR54_SULSO	Q9hnm5 halobacteri
27	693	30.5	456	1 SR54_HALN1	O9hkt0 thermoplasm
28	680	29.9	450	1 SR54_THEAC	Q979y8 thermoplasm
29	667	29.3	455	1 SR54_THEYO	Q9yb62 aeropyrum p
30	664.5	29.2	441	1 SR54_AERPE	P49967 arabidopsis
31	630.5	27.7	495	1 SR53_ARATH	P49972 lycopersico
32	628.5	27.6	499	1 SR52_LYCES	P20424 saccharomyc
33	628.5	27.6	541	1 SR54_YEAST	

## RESULT 1

ID	SR54_BACSU	STANDARD;	PRT;	446 AA.
AC	P37105,			
DT	01-OCT-1994	(Rel. 30, Created)		
DT	01-OCT-1994	(Rel. 30, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Signal recognition particle protein (Fifty-four homolog).			
GN	FFH			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=93338695; PubMed=8335643;			
RT	Honda K., Nakamura K., Nishiguchi M., Yamane K.;			
RT	"Cloning and characterization of a Bacillus subtilis gene encoding a			
RT	homolog of the 54-kilodalton subunit of mammalian signal recognition			
RT	particle and Escherichia coli Ffh."			
RL	J. Bacteriol. 175:4885-4894(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschl C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			
RA	Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,			
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;			
RT	"The complete sequence of the Gram-positive bacterium Bacillus			
RT	subtilis."			
RL	Nature 390:249-256(1997).			

## ALIGNMENTS

```

RN [3]
RP CHARACTERIZATION.
RX MEDLINE=94197732; PubMed=7511896;
RA Nakamura K., Nishiguchi M., Honda K., Yamane K.;
RT "The Bacillus subtilis SRP54 homologue, Ffh, has an intrinsic GTPase
RT activity and forms a ribonucleoprotein complex with small cytoplasmic
RT RNA in vivo.";
RL Biochem. Biophys. Res. Commun. 199;1394-1399(1994).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES.
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL
CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D14356; BA21691.1; -.
CC DR EMBL; Z99112; CAB13471.1; -.
CC DR PIR; B47154; B47154.
CC DR HSPSP; 007347; 1FFH.
CC DR Subtilist; BG10830; ffh.
CC DR InterPro; IPR003593; AAA_AtPase.
CC DR InterPro; IPR000897; SRP54.
CC DR InterPro; IPR004125; SRP54_SPB.
CC DR InterPro; IPR004780; SRP_sub.
CC DR Pfam; PF00448; SRP54; 1.
CC DR Pfam; PF02881; SRP54_N; 1.
CC DR Pfam; PF02978; SRP_SPB; 1.
CC DR ProDom; PD000819; SRP54; 1.
CC DR SMART; SM00382; AAA; 1.
CC DR TIGRFAMS; TIGR00959; 3a05010s1; 1.
CC DR PROSITE; PS00300; SRP54; 1.
CC DR SIGNAL recognition particle; GTP-binding; RNA-binding; Membrane;
CC KW Complete proteome.
CC FT DOMAIN 1 296 G-DOMAIN (BY SIMILARITY).
CC FT DOMAIN 297 446 M-DOMAIN (BY SIMILARITY).
CC FT NP_BIND 108 115 GTP (BY SIMILARITY).
CC FT NP_BIND 191 195 GTP (BY SIMILARITY).
CC FT NP_BIND 249 252 GTP (BY SIMILARITY).
CC SQ SEQUENCE 446 AA; 49541 MW; D71943B23374633D CRC64;

Query Match 72.4%; Score 1647; DB 1; Length 446;
Best Local Similarity 70.1%; Pred. No. 2,le-78;
Matches 319; Conservative 66; Mismatches 60; Indels 10; Gaps 1;

QY 1 MAFEGLSERLOATMOKRGKLEADIKIMREVRALFLADYVNFVKVEFKTVSERA 60
DB 1 MAFEGSLADRLQCTISKRGKGVSEQDVKEMREVRALLADYVNFVKVDSERA 60
QY 61 LGSVDVMSLTGQGVQIVQDELTKLNGENTSNKSNKPTVYVMVYLOGAGKTTTACK 120
DB 61 VGDVMSKSTPGQGVQIVQDELTKLNGENTSNKSNKPTVYVMVYLOGAGKTTTSGK 120
QY 121 LALLMRKKNKPMVAADIRYPAAINQLQTVGKIDIPVYSEGQVQKQOIVTNALKHA 180
DB 121 LANLRRKKNRPMVAADIRYPAIKQLTGLGQVSLGQVSPVEIAKQALEKA 180
QY 181 KEEDHLDVIDTAGRLHIDELALMELKEVKEIAKPNELMYVDSMTGQADVNVASFDDQ 240
DB 181 KEEDHLDVIDTAGRLHIDELALMELKEVKEIAKPNELMYVDSMTGQADVNVASFDDQ 240
QY 241 LDVTGTVTLNKGDTFGGAALSIRSVQKFIKPVGNSEKLDGLELFFPHERMASRILNGD 300

```

```

Db 241 LGLTGVTLNKGDTFGGAALSIRAVTNTPIKFAIGCEKLDALFFPHERMASRILNGD 300
QY 301 YLSLIERAAQDVQVEKAKDLEKKMRESSFTLDDFLEQLDQVKNLGLDDIMKMTIPGNKM 360
Db 301 VLTLLERAAQSDVEDKAKLEQKMTMSFTLDDFLEQLDQVKNLGLDDIMKMTIPGNKM 360
QY 361 KGLDKLNNSEKQIDHAKAIQSMTPAERNPDFTLVSRKKRIAKSGRSLOEYVRLMKQF 420
Db 361 KGLKNIQVDEKQLNHVEALIKMTVLEKEQDPIINASRRRIAKSGTSVQEVNRLKQF 420
QY 421 NDMMKMKQFTGGGKGKGRKNQNMMLKGMNLPF 455
Db 421 DEMKKMKQNTNNSKGGK-----KGFKLPF 445

RESULT 2
SR54_STRMU
ID SR54_STRMU STANDARD; PRT; 516 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (Fifty-four homolog).
GN FFH.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RC STRAIN=JHI1005;
RA Gutierrez J.A., Cvitkovitch D.G., Brady L.J., Hamilton I.R.,
RA Hillman J.D., Blewels A.S.;
RT "Ffh of Streptococcus mutans is involved in acidurance.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RC STRAIN=JHI1005;
RC STRAIN=JHI1005;
RC MEDLINE=96312359; PubMed=8763945;
RA Gutierrez J.A., Crowley P.J., Brown D.P., Hillman J.D., Youngman P.,
RA Blewels A.S.;
RT "Insertional mutagenesis and recovery of interrupted genes of
RT Streptococcus mutans by using transposon Tn917: preliminary
RT characterization of mutants displaying acid sensitivity and
RT nutritional requirements.";
RT J. Bacteriol. 178:4166-4175(1996).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES.
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL
CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH.
CC -1- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U88582; AAB48050.1; -.
CC EMBL; U48983; AAC44500.1; -.
CC HSPSP; 007347; 2FFH.
CC InterPro; IPR000897; SRP54.
CC InterPro; IPR004125; SRP54_SPB.
CC InterPro; IPR004780; SRP_sub.
CC Pfam; PF00448; SRP54; 1.
CC Pfam; PF02681; SRP54_N; 1.

```









```
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74697; CAA98978.1; -.
DR EMBL; AE007120; AAK47310.1; -.
DR HSP; 007347; 2FFH.
DR TIGR; MT2984; -.
DR Tuberculist; Rv2916c; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 301 G-DOMAIN (BY SIMILARITY).
FT NP_BIND 302 525 M-DOMAIN (BY SIMILARITY).
FT NP_BIND 107 114 GTP (BY SIMILARITY).
FT NP_BIND 196 200 GTP (BY SIMILARITY).
FT NP_BIND 254 257 GTP (BY SIMILARITY).
SQ SEQUENCE 525 AA; 55001 MW; 36F6227BC367817C CRC64;

Query Match 48.0%; Score 1091.5; DB 1; Length 525;
Best Local Similarity 47.0%; Pred. No. 1.1e-49;
Matches 224; Conservative 88; Mismatches 132; Indels 33; Gaps 6;

QY 3 FEGLSERLATQWKRGKGIKTEADIKIMRREVRALFADVNFVKKVKEFTKTVSERALG 62
DQ ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
DQ 2 FESLSRLTAALQGLRGRLTDADITTEIRLALLLEADVSLVPVAFTHRIKERARG 61
QY 63 SDVMQSLTFGQGVKTVDETKLMGGENTSNKSNKPTVMYVGLGAGKTTTAGKLA 122
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 62 AEVSSALNFAQQVVKIVNEELISLIGGETRELAFATPTTVVNLAGLQSGKTTLAGKLA 121
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 123 LLMRKYNKPKMLVRAADYRPAAINQLQTVGKQIDIPVYS-----EGDQVRPQQIVTN 175
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 122 ARLGQ-GHTPLVACDLQRPAAVNLQVVGGERAGVPFAPHPGASPEGSPGDPVAVAA 180
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 176 ALKAKEEHLDFVIDTGRHIDEALMNLKEVKEIAPNEMILVDSMTGQDAVNVAE 235
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 181 GLAEARAKHPDVVYVDTAGRIGIDEELMAQAARIRDAINPDEVFLVDAMIGQDAVTAA 240
QY 236 SFDDQLDVTGVTITKLDGTRGGAALSTRVTKPIKFKVGMSEKLDGLLELPEPMASRI 295
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 241 AFGEGVGTGVALTKLDGARGGAALSREVTGVPILFASSTGEKLEDFDFVFPDMASRI 300
QY 296 LGMGDVLSLLEKAQQDDQDOEKADLEKKMRRESSFTLDDLEQLDQVKNLGLPLDIMKMP 355
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 301 LGMGDVLSLLEQAEQVFDQAQAEAAKIGAGELTLEDFLEQLAVRKMGPIGNLLGMLP 360
QY 356 GNNKMGGLDKL-NMSEKQIDHDKALIQSMTFAENNPFLNYSRKRIAKSGSRLQEVN 414
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 361 GAAQMK--DALAEVDKQLDRVQALIRGWTPOERADPKTIINASRLRIANGSGVTSEVN 418
QY 415 RLKQFNDMKKMKQFTG-----GGKGGKGRKQMQNQLKGNLPL 454
DQ :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 419 QLVREFFFEARKMSSMLGGMGIPGIGRKSATRKSGKAGKSGKSK-----KGTRGP 470
```

```
RESULT 8
SR54_MYCLE
ID SR54_MYCLE STANDARD; PRT; 521 AA.
AC O330J3;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (fifty-four homolog).
GN FFH OR ML1622 OR MLCB250.20.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RX Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL
CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH (BY SIMILARITY).
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE. THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z97369; CAB10614.1; -.
DR EMBL; AL583922; CAC30573.1; -.
DR HSP; 007347; 2FFH.
DR Leproma; ML1622; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; FALSE_NEG.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 301 G-DOMAIN (BY SIMILARITY).
FT DOMAIN 302 521 M-DOMAIN (BY SIMILARITY).
FT NP_BIND 107 114 GTP (BY SIMILARITY).
FT NP_BIND 196 200 GTP (BY SIMILARITY).
FT NP_BIND 254 257 GTP (BY SIMILARITY).
SQ SEQUENCE 521 AA; 55111 MW; 37F67B8FC61BDA8B CRC64;

Query Match 47.0%; Score 1069; DB 1; Length 521;
Best Local Similarity 45.2%; Pred. No. 1.6e-48;
Matches 222; Conservative 91; Mismatches 134; Indels 44; Gaps 6;
```





```
Db 78 FGLTGLGLAANSLKGEVLTNDTAEPMDIRRALLEADVSLPVVRFFVQSVSDQAVG 137
QY 63 SDVMOSLPGQGVKIVODELTKLMGENTSNMKNKPTTVMVMVGLQGAGKTTTAGKIA 122
Db 138 MGIVRGVPPDQOLVKIVHDELVLKMGGEVSELQFAGSKGPTVILLAGLQGVGKTTVCAKLA 197
QY 123 LLMRKYNKKPMLVAADYIPAAINLOIVGKQIDIPYVSEGDQVQKQIIVTNALHAKHE 182
Db 198 CYL-KKQSGKSMLIAGVYPAALDQVLVLGEGVGPVYTAGTDVXKPADIAKQGLKEAKK 256
QY 183 EHLDFVIIITAGRLIDHALMNEKVEKIAKPNEIMLVDSMTGQDVAVNAESFDOLD 242
Db 257 NNVDVIMDTAGLQIDKGMDELKQVKKFLAPTEVLLVVDAMTQGAALVYTFNVEIG 316
QY 243 VTGVTTIKLDGTRGGAALSIRSVTKPIKFGVMSEKLDGLFHPHPPERMASRILGMDYL 302
Db 317 ITGAILTKLDGSRGGAALSVEYSGPKIKVGRGERMEDLEPTYPDMAGRILGMDYL 376
QY 303 SLTEKAQDVDOEKAKDLEKXKRESSFTLDDFLEQLDQVKNLGPLDDIMKMIKPMNKKG 362
Db 377 SFVEKATEYMRQEDAEEDLQKKIMSAKFPDNDFLKQTRAVAKMGSMTRVLGXIFGMDGVSP 436
QY 363 LDKLNMSEKQIDHIKAIQSMTPAERNPDNLNVS--RKRRTAKSGRSLOEVNRLMKOF 420
Db 437 A-QIREAKNLLVMEAMLEVMTPEERPELLAESPEREKRIAKDSKGTQEQVYSAVQAI 495
QY 421 NDMKKMMKQFTGGKG 436
Db 496 FQMRVKKNLGMVMEG 511

RESULT 10
SR54_AQUAE STANDARD; PRT; 454 AA.
AC O67615;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (fifty-four homolog).
GN PFH OR AQ.1720.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC MOLECULE AND PROTEIN PFH (BY SIMILARITY).
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; AE000753; AAC07579.1; -.
DR HSP; O07347; lFEH.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 289 G-DOMAIN.
FT DOMAIN 290 454 M-DOMAIN.
FT NP_BIND 102 109 GTP (BY SIMILARITY).
FT NP_BIND 184 188 GTP (BY SIMILARITY).
FT NP_BIND 242 245 GTP (BY SIMILARITY).
SQ SEQUENCE 454 AA; 50911 MW; 9f13fA8C883A0558 CRC64;

Query Match 44.4%; Score 1009; DB 1; Length 454;
Best Local Similarity 45.0%; Pred. No. 1.7e-45;
Matches 194; Conservative 103; Mismatches 8; Gaps 3;

QY 4 EGLSEKLTQTMKMGKGLTEADIKIMREVRLALFADVNFVKVEFKITYSERALGS 63
Db 3 ELUTEKFSHALEKLTNARKITEKNINQTLREIFLLEADVQYAKDFIKRIKRVKVG 62
QY 64 DVMQSLTPGQGVKIVODELTKLMGENTSNMKNKPTTVMVMVGLQGAGKTTAGKIAL 123
Db 63 EYFANLSPREFVIKYYEELVDILGGEKADLKKG-----TVLFVGLQGTSTTTISKIAN 117
QY 124 LMRKYNKKPMLVAADYIPAAINLOIVGKQIDIPVYSEGDQVQKQIIVTNALKHAKEE 183
Db 118 LL-KKGHKVAVSVSTDLRRPAAMLQQLAERAVGVYPYEFEGGLGAVEIARRAVKRAKEE 176
QY 184 HLDFVIIDTAGRLIHDEALMNEKVEKIAKPNEIMLVDSMTGQDVAVNAESFDOLDV 243
Db 177 SVDYLLDITAGRLHYDELMKELQEKIEKVTNPSEIILYVADAMOGGTALETAKTFHERLGL 236
QY 244 TGYTLFKLDGTRGGAALSIRSVTKPIKFGVMSEKLDGLFHPHPPERMASRILGMDGVLS 303
Db 237 TGVYTKMGDARGGLALSVEKVLGVPIKFIGVEKIEDIEFFYPDRIAQRLGLGDIQS 296
QY 304 LIEKAQDVDOEKAKDLEKXKRESSFTLDDFLEQLDQVKNLGPLDDIMKMIKPMNKKGL 363
Db 297 LVEKAQEVITDKAQVMAKVMKTGTFDLEDLEMLRMIOQMGPDLKLLSMIPGVAPO--L 354
QY 364 DKLNMSEKQIDHIKAIQSMTPAERNPDNLNVSRRKKRIAKSGRSLOEVNRLMKQFNDM 423
Db 355 KHLKVDKQKQFKIEALINSMTPEERNRPNKIINSRKKRIARGSGTIVSDVKNLLKRYEEM 414
QY 424 KMMKMQFTGGG 434
Db 415 KMMRKLQKAG 425

RESULT 11
SR54_BUCAI STANDARD; PRT; 451 AA.
ID P57473; O914J1;
AC P57473; O914J1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (fifty-four homolog).
GN PFH OR BU393.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
```



```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR EMBL; U82109; AAB58502.1; -.  
DR PDB; 1FFH; 31-DEC-97.  
DR PDB; 2FFH; 16-JUL-99.  
DR PDB; 1NG1; 13-MAR-00.  
DR PDB; 2NG1; 13-MAR-00.  
DR PDB; 3NG1; 13-MAR-00.  
DR InterPro; IPR003593; AAA_ATPase.  
DR InterPro; IPR000897; SRP54.  
DR InterPro; IPR004125; SRP54_SPB.  
DR InterPro; IPR004780; SRP_sub.  
DR Pfam; PF00448; SRP54; 1.  
DR Pfam; PF02881; SRP54_N; 1.  
DR Pfam; PF02978; SRP_SPB; 1.  
DR ProDom; PD000819; SRP54; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.  
DR PROSITE; PS00300; SRP54; 1.  
KW Signal recognition particle; GTP-binding; RNA-binding; 3D-structure.  
FT INIT_MET 0 0  
FT DOMAIN 1 291 G-DOMAIN.  
FT DOMAIN 292 429 M-DOMAIN.  
FT NP_BIND 104 111 GTP (BY SIMILARITY).  
FT NP_BIND 186 190 GTP (BY SIMILARITY).  
FT NP_BIND 244 247 GTP (BY SIMILARITY).  
SQ SEQUENCE 429 AA; 47225 MW; 26D387A5A303EFBD CRC64;  
  
Query Match 44.2%; Score 1005.5; DB 1; Length 429;  
Best Local Similarity 47.5%; Pred. No. 2.4e-45;  
Matches 209; Conservative 92; Mismatches 118; Indels 21; Gaps 6;  
  
QY 3 FEGLSERLQATMOKMGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERALG 62  
D 1 FQOLSARLQEAIGRLGRGRITEEDLKATREIRRALMDADVNLEAYDFVERVREBALG 60  
  
QY 63 SDVMQSLTPGQGVKIVQDELTKLMGENTSIINMSKNPPTVMVGLQAGKTTACKLA 122  
D 61 KOVLESUTPAEVLIAATVVEALKEALGGEARLPVLKDR--NLWFLVGLQSGKTTAAKLA 118  
  
QY 123 LLMKKYKKNKPMPLVAADIYRPAAINQLOTGVKQIDIPVSEGDQVKKPQOIVTNALKHAKE 182  
D 119 LYYKKG-GRPLVLAADTORPAAREQLRLGKGVVPLEYMDGESPSIRRYVEKARL 177  
  
QY 183 EHLDFVIDTAGRLHIDEALMNEKVEKIAKNEIMLVVDSMTGQDVAVNAESFDQOLD 242  
D 178 EARDLILVDTAGRLQIDPLMGEARLKEVLGPDVLLVLDAMTGOEALSVARAFDEKVG 237  
  
QY 243 VTGVLTKLDGTRGGAALSIRSVTKPKIFGVGSEKLDGLELHPERWASRIILGMGDLV 302  
D 238 VTGVLVTKLDGARGGAALSARHVTGKPIYFAGVSEKPEGLEPFPERLAGRIILMGDVA 297  
  
QY 303 SLIEKAQDDVQEKADLEKKMKRESS--FTLDDFLEQLDQVKMLAGPLDDIKMIPGNKMK 360  
D 298 SLAEKV-----RAAGLEAREAPKSAKELSLEDLQKQNLKELGFFSEILGLPGVPQ- 349  
  
QY 361 KGLDKLMSKQIDHKAIIQSMTPAERNPDTLNVRKRIAKSGRSLSQEVNRLMKQF 420  
D 350 ---GLKVDKAIKLEAVLSMTPEERKDPRIILNGSRKRIAKSGTSVQEVNRFKAF 405  
  
QY 421 NDMKMMKQFTGGGKKKK 440  
D 406 EEMKALMSL-----EKKKGR 421  
  
RESULT 13  
ID SR54_RICPR STANDARD; PRY; 449 AA.  
AC Q96D20;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE Signal recognition particle protein (Fifty-four homolog).  
GN FFH OR RPL73.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsieae; Rickettsia.  
OX NCBL_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Madrid E;  
RA MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria";  
RL Nature 396:133-140(1998).  
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC  
CC -!- PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE  
CC RIBOSOMES (BY SIMILARITY).  
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA  
CC MOLECULE AND PROTEIN FFH (BY SIMILARITY).  
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN  
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL  
CC SEQUENCE.  
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR EMBL; AJ235270; CAA14640.1; -.  
DR HSSP; C07347; 1FFH.  
DR InterPro; IPR000897; SRP54.  
DR InterPro; IPR004125; SRP54_SPB.  
DR InterPro; IPR004780; SRP_sub.  
DR Pfam; PF00448; SRP54; 1.  
DR Pfam; PF02881; SRP54_N; 1.  
DR Pfam; PF02978; SRP_SPB; 1.  
DR ProDom; PD000819; SRP54; 1.  
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.  
DR PROSITE; PS00300; SRP54; 1.  
KW Signal recognition particle; GTP-binding; RNA-binding;  
KW Complete proteome.  
FT DOMAIN 1 296 G-DOMAIN.  
FT DOMAIN 297 449 M-DOMAIN.  
FT NP_BIND 109 116 GTP (BY SIMILARITY).  
FT NP_BIND 191 195 GTP (BY SIMILARITY).  
FT NP_BIND 249 252 GTP (BY SIMILARITY).  
SQ SEQUENCE 449 AA; 49279 MW; 6D80817746BFA045 CRC64;  
  
Query Match 42.2%; Score 958.5; DB 1; Length 449;  
Best Local Similarity 44.2%; Pred. No. 6.8e-43;  
Matches 200; Conservative 99; Mismatches 135; Indels 19; Gaps 8;  
  
QY 3 FEGLSERLQATMOKMGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERALG 62  
D 2 FXTLTQNLTKIFDKLYNSGILTFENQIDTAMRDVRVALLESVALPVKGFIEVKQKALG 61  
  
QY 63 SDVMQSLTPGQGVKIVQDELTKLMG--ENTSINMSKNPPTVMVGLQAGKTTAGK 120  
D 62 QEVKSVSPGMIITITHEEMINLLASTESTTKLNNAKPPVNLIVLQGGKTTASGK 121  
  
QY 121 LALLMRKYNKKPMPLVAADIYRPAAINQLOTGVKQIDI---PVYSEGQVKEPQIVNAL 177  
D 122 LALRL-KNQNKVLLVSLTYRPAAQEQALAIANSVNSLPIV-KGE--KPLDIVKRAI 177  
  
QY 178 KHAKHEHLDFVIDTAGRLHIDEALMNEKVEKIAKNEIMLVVDSMTGQDVAVNAESF 237  
D 178 GEAQISADVVIDTAGRTQIDNVNMEALAKKILNPTETLLVIDSMTGQDAVITAKTF 237
```



```
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39684; AAC71264.1; -.
DR HSP; O07347; 1FTH.
DR TIGR; MG048; -.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR TrEMBL; TIGR00959; 3a0501sol; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 293 G-DOMAIN.
FT DOMAIN 294 446 M-DOMAIN.
FT NP_BIND 106 113 GTP (BY SIMILARITY).
FT NP_BIND 188 192 GTP (BY SIMILARITY).
FT NP_BIND 246 249 GTP (BY SIMILARITY).
SQ SEQUENCE 446 AA; 50199 MW; 080F225C8B5E5AE CRC64;

Query Match 39.5%; Score 898; DB 1; Length 446;
Best Local Similarity 43.0%; Pred. No. 8.8e-40;
Matches 198; Conservative 90; Mismatches 135; Indels 38; Gaps 10;

QY 6 LSERLQATQMKRKGKTEADIKIMREYRLALFEADYVKNVKEIKTVSERALGSDV 65
DB 6 LSSIVMTQKKNIAQTTEKDVLEVLKEIRIALDLADVNLVKNFIKAIRDYGVG--- 62

QY 66 MQSLTPGQ---QVIKIVQDELTKMGENTSINMSNKPPTVYMMVGLQGAGKTTAGKL 121
DB 63 -QTEPGQDLQKSLKTKTELINLSQPNQELN--EKRLKIMVGLQSGSKTTTCGKL 119

QY 122 ALLMRKYNKKPMLVAADIIYPAAINOLQVKGQIDIPVYSEGQ--VKPQIIVTNALKH 179
DB 120 AYWLKRYKQKTMVLGDIYRPAATGLETLSQQTNSVFFAQGTQPVAKTTKAALSAFKT 179

QY 180 AKESHLDFVIIDTAGRLHIDEALMNEKVEKTAKEIMLYVDSMTGQDAVNVASFEED 239
DB 180 AK--COTIICDTAGRLQTNELMDELVSVKNEINPEDEIMVVDLSQGEIINVAQTFHK 236

QY 240 QLDVGTGTLTKDGTGRGAALSIRSYTQPKIFVGMSEKLDGLDFHPERMASRIILGMG 299
DB 237 RLKLTGFIISKLDSDARAGAALSASLLQVPKILGVSEKLDGLEQHPERIANRIILGLG 296

QY 300 DVLSLETKAQDDVDQEKANDLEK--KWRSSSTLDDFLDQLDVKNLGLPDDIMKNMIPG 356
DB 297 DVMSLVEKABQVFDK---KDLTKTISKMFLGKMDLEDLLIYMQQMHKMGVSSLIKMLPA 353

QY 357 MNKMKGLDKLNSEKQIDHIAIKAIQSTPAERNPDITN--VSRKKRIAKSGHSLQEVN 414
DB 354 -NFSVSENAELIENKIELVNLINSMTREERHPKLIINRDPNRRKRIKSGSKMDLNL 412

QY 415 RLMQFNDMKMKMKQFTGGGKKGKKNQNMKGMNLPLF 455
DB 413 KLMKFNKML-----KATEMKLLKTKGSNPF 439
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:15:12 ; Search time 35 Seconds  
(without alignments)  
2678.612 Million cell updates/sec

Title: US-09-943-108A-2  
Perfect score: 2274  
Sequence: 1 MAFESGLSERLQATMQRKQK.....GKKGKRQNMQLKGNLPLF 455

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB ID	Description
1	2274	100.0	455	16 Q99UN3	Q99UN3 staphylococ
2	1717.5	75.5	450	16 Q8Y695	Q8Y695 listeria mo
3	1708.5	75.1	450	16 Q92AK7	Q92AK7 listeria in
4	1584	69.7	451	16 Q9KA10	Q9KA10 bacillus ha
5	1393	61.3	449	16 Q97I98	Q97I98 clostridium
6	1382	60.8	452	16 Q8XJP3	Q8XJP3 clostridium
7	1352.5	59.5	447	16 Q8R9X0	Q8R9X0 thermococ
8	1315	57.8	518	16 Q9CF65	Q9CF65 lactococcus
9	1306	57.4	523	16 Q97OD2	Q97OD2 streptococ
10	1279.5	56.3	519	16 Q98ZK1	Q98ZK1 streptococ
11	1238	54.4	444	16 Q8RDV7	Q8RDV7 fusobacteri
12	1180.5	51.9	433	16 Q9XIQ1	Q9XIQ1 thermotoga
13	1173.5	51.6	453	16 Q8XF48	Q8XF48 salmonella
14	1150.5	50.6	461	16 Q9KUG1	Q9KUG1 vibrio chol
15	1150	50.6	457	16 Q9HXP8	Q9HXP8 pseudomonas
16	1139.5	50.1	453	16 Q8ZBU6	Q8ZBU6 yersinia pe

17	1135.5	49.9	458	16 Q9CLN6	Q9CLN6 pasteurella
18	1131	49.7	456	16 Q9K177	Q9K177 neisseria m
19	1130	49.7	456	2 Q9JPK6	Q9JPK6 neisseria m
20	1130	49.7	456	16 Q9USQ0	Q9USQ0 neisseria g
21	1123	49.4	456	2 Q8RNK2	Q8RNK2 neisseria g
22	1121.5	49.3	458	2 Q9L9C9	Q9L9C9 thiobacillu
23	1099.5	48.4	476	16 Q8XVL9	Q8XVL9 ralstonia s
24	1092.5	48.0	490	16 Q8VVM3	Q8VVM3 anabaena sp
25	1081.5	47.6	532	16 Q98E72	Q98E72 rhizobium i
26	1077.5	47.4	550	16 Q69874	Q69874 streptomyce
27	1070.5	47.1	550	2 Q85676	Q85676 streptomyce
28	1066.5	46.9	452	16 Q9RTC9	Q9RTC9 deinococcus
29	1062.5	46.7	554	16 Q8UC01	Q8UC01 agrobacteri
30	1054	46.4	523	16 Q8YJ61	Q8YJ61 brucella me
31	1050.5	46.2	449	16 Q98QL9	Q98QL9 mycoplasma
32	1050.5	46.2	513	16 Q92L45	Q92L45 rhizobium m
33	1039	45.7	513	2 Q8VVP8	Q8VVP8 arcanobacte
34	1018	44.8	434	16 Q9PH73	Q9PH73 xylella fas
35	981.5	43.2	449	16 Q92J55	Q92J55 rickettsia
36	981	43.1	525	10 Q82532	Q82532 pisum sativ
37	970.5	42.7	508	16 Q9A2B3	Q9A2B3 caulobacter
38	933.5	41.1	448	16 Q9PR53	Q9PR53 ureaplasma
39	905.5	39.8	447	16 Q51637	Q51637 borrelia bu
40	899	39.5	549	10 Q9AXU1	Q9AXU1 chlamydomon
41	888	39.1	448	16 Q9PL14	Q9PL14 chlamydia m
42	869	38.2	448	16 Q84028	Q84028 chlamydia t
43	846	37.2	448	16 Q9JRV9	Q9JRV9 chlamydia p
44	846	37.2	448	16 Q9Z966	Q9Z966 chlamydia p
45	843	37.1	445	16 Q9PPJ8	Q9PPJ8 campylobact

ALIGNMENTS

RESULT 1

Q99UN3 PRELIMINARY; PRT; 455 AA.

AC Q99UN3; 2001 (TREMREL. 17, Created)

DT 01-JUN-2001 (TREMREL. 17, Last sequence update)

DT 01-JUN-2002 (TREMREL. 21, Last annotation update)

DE Ffh protein (Signal recognition particle homolog).

GN Ffh OR SAV1237 OR SA1080.

OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and

OS Staphylococcus aureus (strain N315).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Staphylococcus.

OX NCBI\_TaxID=158879; 158879;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);

RX MEDLINE=21311952; PubMed=11418146;

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,

RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,

RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,

RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,

RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,

RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;

RT "Whole genome sequencing of methicillin-resistant Staphylococcus aureus."

RL Lancet 357:1225-1240(2001).

DR EMBL; AP003361; BAB57399.1; -.

DR EMBL; AP003133; BAB42332.1; -.

DR HSSP; O07347; 1FFH.

DR InterPro; IPR003593; AAA\_ATPase.

DR InterPro; IPR000897; SRP54.

DR InterPro; IPR004125; SRP54\_SBP.

DR InterPro; IPR004780; SRP\_sub.

DR Pfam; PF00448; SRP54; 1.

DR Pfam; PF02881; SRP54\_N; 1.

DR Pfam; PF02978; SRP\_SBP; 1.

DR ProDom; PD000819; SRP54; 1.

```
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Complete proteome.
SQ SEQUENCE 455 AA; 50753 MW; 619A202B1734097C CRC64;

Query Match
Best Local Similarity 100.0%; Score 2274; DB 16; Length 455;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSNMKNPPTVVMVGLQGAGKTTTGGK 120
DB 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSNMKNPPTVVMVGLQGAGKTTTGGK 120
QY 121 LALLMRKYNKPMVAADIYRPAALNQLQTVGKQIDIPYSEGQDVQKQIIVTNALKHA 180
DB 121 LALLMRKYNKPMVAADIYRPAALNQLQTVGKQIDIPYSEGQDVQKQIIVTNALKHA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKETAKPNEIMLVDSMTGQDAVNVASFDDQ 240
DB 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKETAKPNEIMLVDSMTGQDAVNVASFDDQ 240
QY 241 LDVTGVTITKLDGTRGGAALSIRSVTKPIKFGVMSSEKLDGLFELFHPERMASRILGMD 300
DB 241 LDVTGVTITKLDGTRGGAALSIRSVTKPIKFGVMSSEKLDGLFELFHPERMASRILGMD 300
QY 301 VLSLIEKAQVDQEKAKOLEKKMRSEFTLDDFLEQLDQVKNLGPLDDIMKPIGMNKM 360
DB 301 VLSLIEKAQVDQEKAKOLEKKMRSEFTLDDFLEQLDQVKNLGPLDDIMKPIGMNKM 360
QY 361 KGLDLNMSKQIDHKAIIQSTPAERNPDTLNVSRKRIAKGSSRSLOEVNRLMKQF 420
DB 361 KGLDLNMSKQIDHKAIIQSTPAERNPDTLNVSRKRIAKGSSRSLOEVNRLMKQF 420
QY 421 NDMKKMKQFTGGGKGGKGRNQMNKLGKMLPF 455
DB 421 NDMKKMKQFTGGGKGGKGRNQMNKLGKMLPF 455

RESULT 2
Q8Y695 PRELIMINARY; PRT; 450 AA.
AC Q8Y695;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ffh protein.
GN Ffh OR LMO1801.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Maifournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voes H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species."
RL Science 294:849-852(2001).
```

```
DR EMBL; AL591981; CAC99879.1; -.
DR ListList; LMO01801; -.
DR InterPro; IPR003593; AAA_AtPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54.N; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
KW Complete proteome.
SQ SEQUENCE 450 AA; 49714 MW; D5888D7A99C60C76 CRC64;

Query Match
Best Local Similarity 75.5%; Score 1717.5; DB 16; Length 450;
Matches 331; Conservative 64; Mismatches 55; Indels 5; Gaps 1;

QY 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSNMKNPPTVVMVGLQGAGKTTTGGK 120
DB 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSNMKNPPTVVMVGLQGAGKTTTGGK 120
QY 121 LALLMRKYNKPMVAADIYRPAALNQLQTVGKQIDIPYSEGQDVQKQIIVTNALKHA 180
DB 121 LALLMRKYNKPMVAADIYRPAALNQLQTVGKQIDIPYSEGQDVQKQIIVTNALKHA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKETAKPNEIMLVDSMTGQDAVNVASFDDQ 240
DB 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKETAKPNEIMLVDSMTGQDAVNVASFDDQ 240
QY 241 LDVTGVTITKLDGTRGGAALSIRSVTKPIKFGVMSSEKLDGLFELFHPERMASRILGMD 300
DB 241 LDVTGVTITKLDGTRGGAALSIRSVTKPIKFGVMSSEKLDGLFELFHPERMASRILGMD 300
QY 301 VLSLIEKAQVDQEKAKOLEKKMRSEFTLDDFLEQLDQVKNLGPLDDIMKPIGMNKM 360
DB 301 VLSLIEKAQVDQEKAKOLEKKMRSEFTLDDFLEQLDQVKNLGPLDDIMKPIGMNKM 360
QY 361 KGLDLNMSKQIDHKAIIQSTPAERNPDTLNVSRKRIAKGSSRSLOEVNRLMKQF 420
DB 361 KGLDLNMSKQIDHKAIIQSTPAERNPDTLNVSRKRIAKGSSRSLOEVNRLMKQF 420
QY 421 NDMKKMKQFTGGGKGGKGRNQMNKLGKMLPF 455
DB 421 NDMKKMKQFTGGGKGGKGRNQMNKLGKMLPF 455

RESULT 3
Q92AK7 PRELIMINARY; PRT; 450 AA.
AC Q92AK7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ffh protein.
GN Ffh OR LIN1915.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
```



```

RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species."
RL Science 294:849-852(2001).
DR EMBL; AL596170; CAC97145.1; -.
DR Listlist; LIN01915; -.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
KW Complete proteome.
SQ SEQUENCE 450 AA; 49684 MW; E417626DID69D239 CRC64;

Query Match 75.1%; Score 1708.5; DB 16; Length 450;
Best Local Similarity 72.3%; Pred. No. 9.7e-84;
Matches 329; Conservative 65; Mismatches 56; Indels 5; Gaps 1;

QY 1 MAPEGLSERLOATMOKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAPEGLSERLOATMOKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSVMOSLTGPGQVQIKIVODELTKLMGENTSIKNSKPPPTVVMVGLQAGKTTTAK 120
DB 61 VGDVWKSLTGPGQVQIKIVODELTKLMGENTSIKNSKPPPTVVMVGLQAGKTTTAK 120
QY 121 LALLMRKYNKPMVAADYIRPAAINQLQTVGKQIDIPVYSEGQVQKQPOIVTNALKHA 180
DB 121 LALLMRKYNKPMVAADYIRPAAINQLQTVGKQIDIPVYSEGQVQKQPOIVTNALKHA 180
QY 181 KEEHLDFVIDTAGRLHIDEALMELKEVEIAKPNEIMLVVDSMTGQDVAVNAESFDDQ 240
DB 181 KEEHLDFVIDTAGRLHIDEALMELKEVEIAKPNEIMLVVDSMTGQDVAVNAESFDDQ 240
QY 241 LDVTGVTLLKLDGTRGGAALSIRSVTKPIKFGVMSEKLDGLFELFHPERMASRIILMGD 300
DB 241 LEITGVVLLKLDGTRGGAALSIRSVTKPIKFGVMSEKLDGLFELFHPERMASRIILMGD 300
QY 301 VLSLIEKAQDVQDEKADLEKKMRESEFTLDDFLQDLQVKNLGPLDDIMKTPGMNKM 360
DB 301 VLSLIEKAQDVQDEKADLEKKMRESEFTLDDFLQDLQVKNLGPLDDIMKTPGMNKM 360
QY 361 KGLDKLNMSEKQIDHKAIIQSMTPAERNPDPLNYSRKKRIAKSGRSLOEVNRLMKQF 420
DB 361 KGLDKLNMSEKQIDHKAIIQSMTPAERNPDPLNYSRKKRIAKSGRSLOEVNRLMKQF 420
QY 421 NDMMKMMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455
DB 421 AEMKMMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455

RESULT 4
Q9KAL0 PRELIMINARY; PRT; 451 AA.
AC Q9KAL0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Signal recognition particle.
GN FFH OR BH2484.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001515; BAB06203.1; -.
DR HSSP; O07347; 1FFH.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Complete proteome.
SQ SEQUENCE 451 AA; 50104 MW; 1B61BEF45E14674F CRC64;

Query Match 69.7%; Score 1584; DB 16; Length 451;
Best Local Similarity 67.3%; Pred. No. 4.4e-77;
Matches 307; Conservative 73; Mismatches 70; Indels 6; Gaps 2;

QY 1 MAPEGLSERLOATMOKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAPEGLSERLOATMOKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSVMOSLTGPGQVQIKIVODELTKLMGENTSIKNSKPPPTVVMVGLQAGKTTTAK 120
DB 61 LGQVWKSLTGPGQVQIKIVODELTKLMGENTSIKNSKPPPTVVMVGLQAGKTTTAK 120
QY 121 LALLMRKYNKPMVAADYIRPAAINQLQTVGKQIDIPVYSEGQVQKQPOIVTNALKHA 180
DB 121 LALLMRKYNKPMVAADYIRPAAINQLQTVGKQIDIPVYSEGQVQKQPOIVTNALKHA 180
QY 181 KEEHLDFVIDTAGRLHIDEALMELKEVEIAKPNEIMLVVDSMTGQDVAVNAESFDDQ 240
DB 181 KEEHLDFVIDTAGRLHIDEALMELKEVEIAKPNEIMLVVDSMTGQDVAVNAESFDDQ 240
QY 241 LDVTGVTLLKLDGTRGGAALSIRSVTKPIKFGVMSEKLDGLFELFHPERMASRIILMGD 300
DB 241 LDVSGVVTLLKLDGTRGGAALSIRSVTKPIKFGVMSEKLDGLFELFHPERMASRIILMGD 300
QY 301 VLSLIEKAQDVQDEKADLEKKMRESEFTLDDFLQDLQVKNLGPLDDIMKTPGMNKM 360
DB 301 VLSLIEKAQDVQDEKADLEKKMRESEFTLDDFLQDLQVKNLGPLDDIMKTPGMNKM 360
QY 361 KGLDKLNMSEKQIDHKAIIQSMTPAERNPDPLNYSRKKRIAKSGRSLOEVNRLMKQF 420
DB 361 KGLDKLNMSEKQIDHKAIIQSMTPAERNPDPLNYSRKKRIAKSGRSLOEVNRLMKQF 420
QY 421 NDMMKMMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455
DB 421 EEMKMMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455

RESULT 5
Q97198 PRELIMINARY; PRT; 449 AA.
AC Q97198;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE Signal recognition particle.
GN CAC1754.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/clostridium group; Clostridia;

```

OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1488;

```

[1]
SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Neolling J., Breton G., Omelchenko M.V., Makarov K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RT J. Bacteriol. 183:4823-4838(2001).
RL EMBL: AE007684; AAK79720.1; -.
DR InterPro; IPR003593; AAA_ATPase
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE; PS00300; SRP54; UNKNOWN_1.
DR Complete proteome.
KW SEQUENCE 449 AA; 49555 MW; FF4695E75A673733 CRC64;
SQ

```

Query Match 61.3%; Score 1393; DB 16; Length 449;  
Best Local Similarity 61.0%; Pred. NO. 6.9e-67;  
Matches 280; Conservative 71; Mismatches 94; Indels 14; Gaps 5;

QY	1	MAFGLSERIQATQMKRGKLTETADIKIMMREVRALFEADVNFVKVVKFIKTVSERA	60
Db	1	MAFGLSTKLQAMKKLRGKLSKIDKAMREKYLALAEADVNTKIYKNFVKVGEKC	60
QY	61	LGSDDVMSLTGPGQVQIKTVQDELTKLWGENTSIWNKSKPTVYVMVGLQAGKTTTAK	120
Db	61	LGNEVMESLTGPGQVQIKIVNEELNLNMGKEESKIFTEENGITVIMAVGLQAGKITTCGK	120
QY	121	LALLMRKKYNKPKMLVAADIIYRPAANLQLOTYVKGOIDIPYSEGDQVFPQOITVNALKHA	180
Db	121	LSLSLKKK-NKRPLLVAODIYRPAATKQLEVYVGSIDVPVFSNGDKVNPVDISKAAMKHA	179
QY	181	KEEHLDPVIDTAGPLHDEALMNELKYEKETAKEINMLWDSMTQDQVNVVAESDDQ	240
Db	180	KENGANVVIIDTAGRLHHDDGLMNEJENIKSEVNPKEILLVDSMTQDQVNVVAESPDNK	239
QY	241	LDVTGVTLTKLDGTRGGAALSTRVQTKPIKFVGMSKLDGLFHPHPKASRIILGMGD	300
Db	240	LELTGVVLTCLKDGTGGAALSTRMTGAPIKYVGLGEKMDIEIFHPDWRASRIILGMGD	299
QY	301	VLSLIEKAQODVDEKADLEKMKRESSLTDDPLEQLDQVKNLGDPLDDIMKPIPGHAK-	359
Db	300	VTLTIEKAQSAIDEXQAKELGRMLSGEENFDFFLQAFQOMKGLPIGKLELWETPGFNSS	359
QY	360	-MKGLKNNKSEKQDHKALIQSMTPAERNPD--TATYSRKRKTAGSGRSLSQEVNRL	416
Db	360	MLKGYD-LSKNEGEMKYEALIKSMTAKERNKPSLITASTSKRRRLTALGCTTQVEYNKI	418
QY	417	MQFNDNMKKMMKQFTGGGKGGKKGRNMOMMLKGNLPDF	455
Db	419	LKNFEQMKMMKQFKGNFKSKG-----LFGCKMPF	449

```
RESULT 6  
Q8XJP3  
ID Q8XJP3 PRELIMINARY; PRT; 452 AA.  
AC AC  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
```

DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Signal recognition particle protein.
GN	FFH OR CPE1713.
OS	Clostridium perfringens
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC	Clostridiales; Clostridiaceae; Clostridium.
NC	NCBI_TaxID=1502;
LN	[1]
RP	SEQUENCE FROM N.A.
EC	STRAIN-13 / TYPE A;
RX	PubMed=11792842;
RA	Shinizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT	"Complete genome sequence of Clostridium perfringens, an anaerobic
RT	flesh-eater";
RL	Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR	EMBL; AP003191; BAB81419.1; -.
DR	InterPro; IPR003593; AAA_ATPase.
DR	InterPro; IPR000897; SRP54.
DR	InterPro; IPR004125; SRP54_SPB.
DR	InterPro; IPR004780; SRP_sub.
DR	Pfam; PF00448; SRP54; 1.
DR	Pfam; PF02881; SRP54_N; 1.
DR	Pfam; PF02978; SRP_SPB; 1.
DR	ProDom; PD000819; SRP54; 1.
DR	SMART; SM00382; AAA; 1
DR	TIGRFAMS; TIGR00959; 3a0501s01; 1.
DR	PROSITE; PS00300; SRP54; UNKNOWN_1.
SW	Complete proteome.
KO	SEQUENCE 452 AA, 49685 MW: 2356112EBC9C7D31 CRC64:

Query Match	60.8%;	Score 1382;	DB 16;	Length 452;
Best Local Similarity	62.7%;	Pred. No. 2.7e-66;		
Matches 279;	Conservative 72;	Mismatches 86;	Indels 8;	Gaps 5;

Qy	1	MAFEGJSERLOATWOKXRGKGLFEADIKIMREVRVLFALFEADVNFVKVPKEPIKTVSERA	60
Dd	1	MAFDGASLKLDQDFKLKLGKGLTEKDIEAMREVKLALLEADVNFVKVPKFISNVKDKC	60
Qy	61	LGSVDNOSLTPGOOVLIKIVQBELTKLMGCENTSINNSNKPTVTVMVLGLOGAGKTTTAGK	120
Dd	61	VGEEVNLSTPGOOVIKVINOVELLTMLGETESKLAYSONGFTFVLVLGLOGAGKTTMAGK	120
Qy	121	LALLMRKYNNKKPMLYAADIYRPAAINOLQTGVGKOIDIPVYSBGDQPKQQIWTNALKHA	180
Dd	121	LALHLRRK-NKKPELLVACDIYRPAIAIKQLVGVGKOIDIPVFSMGDKVAKAVDIAKAATEHA	179
Qy	181	KEEHLDFVIDTAGRLHIDEALMNELKEYKETAKPNEIMLVYDSMTGQDAVNVAESPDDQ	240
Dd	180	KONGNNVVIIITAGRULHIDELMQELDKVKNSEPEILLVVUAMTGGQAVNVAETFNNS	239
Qy	241	LDTVGTVIIITKLDGTRGGAALSIRSVTQKPIKFVGMSSEKLDGLELPHFERMASRTIGMGD	300
Dd	240	LDLSGIITILTKLDGTRGGAALSIRDITGKPIKFVGVEKMSDIEVFHPDRMASRTIGMGD	299
Qy	301	VUSLIERAQODVOEKARDLEKKPRESSTFLDDFLEQLDQVKNLGPLDDIMKAPGMN--	358
Dd	300	VUSLIERAQAIDDQDEASKLERMLNGEFNFDDYLSAMDQMKLIGPINKLIELMITPGVNTK	359
Qy	359	KMKGLDKLNSEKQIDHKAIIOQSMTFAERNFPDTL--NVSRKKRIKAGSGRSLOEQVNRIL	416
Dd	360	ELEGID-FSOGEKQMATVKAIIQSMTAKEROPSILVICNGSRKRRLAKGSGTITVQENVKV	418
Qy	417	MKQFNDDMMKMFOFGGKG--GGKG 439	
Dd	419	LKGYEMKKQMCKQSFORKNAKSG 443	

```

RESULT 7
Q8R9X0
ID Q8R9X0 PRELIMINARY; PRT; 447 AA
AC Q8R9X0;
BT 01-JUN-2002 (TREMBUREL, 21, Created)

```

01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
Signal recognition particle GTPase.  
FFH OR TrE1462.  
Thermoanaerobacter tengcongensis.  
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.  
NCBI\_TaxID=119072;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=MB4T / JCM11007;  
RX MEDLINE=21992816; PubMed=11997336;  
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
Tan H., Chen R., Wang J., Yu J., Yang H.;  
RT "A complete sequence of T. tengcongensis genome.";  
RL Genome Res. 12:689-700(2002).  
DR EMBL; AE013104; AAM24684.1; -.  
KW Complete proteome.  
SQ SEQUENCE 447 AA; 49294 MW; 3DB00C468FA14B7E CRC64;  
  
Query Match 59.5%; Score 1352.5; DB 16; Length 447;  
Best Local Similarity 59.9%; Pred. No. 1e-64;  
Matches 273; Conservative 74; Mismatches 96; Indels 13; Gaps 4;  
  
QY 1 MAFEGLSERLOATMOKMRGKGLTEADIKIMREVRLEALFEADVNFVKVKEFIKTVSERA 60  
DB 1 MAFESLSERLOQVFKLRGKGLTEKDEKAMREVKVALLERDVNFVKVGFINSVTEKA 61  
  
QY 61 LGSVMQSLTPGGQVQIKIVODELTKLGGENTSNMNSKPPTVYMMVGLQGAGKTTAGK 120  
DB 62 LGQVMSLSRPAQOVQIKIVHEELIKLGSVESRLNLGSKVPVIMMVGLOGSGKTTACGK 121  
  
QY 121 LALLMRKYNKPKMLVAADYIPRAINOLOTVGQIDIPVYSEGQVKKPQIVTNALKHA 180  
DB 122 LANLL-KKQGNPLLVACDTPVPAIKQLVLGANINVPVFTMGDKVDTADIAKASDYA 180  
  
QY 181 KEHLDFVIIDTAGRLHIDEALMNELEKVEIAKPNELMVDVMTGQDVNVAESFDDQ 240  
DB 181 KSHNVVDVIIDTAGRLHIDEALMNELEKVEIAKPNELMVDVMTGQDVNVAESFDDQ 240  
  
QY 241 LDVGTVLTLDGDTGGAALSIRSVTKPIKVFVGMSEKLDGLFELPHPERMASRILMGD 300  
DB 241 LDITGVLTLDGDTGGAALSIRSVTKPIKVFVGMSEKLDGLFELPHPERMASRILMGD 300  
  
QY 301 VLSLIEKAQDVQDEKAKOLEKKMRESSFTLDDFLEQLDQVKNLGPLDDIMKMPGNMK 360  
DB 301 VLTIEKAQADIDEKAKALEMGKILSKOFTLEDFLEQLSLKNMGLPQLLAMIPGVNK- 359  
  
QY 361 KGLDLKLNSEKQIDHIKAIQSWTPAERNPDTLNVRKKRIAKGSGRSLQEVNRLMKQF 420  
DB 360 SVLNVSEKDLKRIEAILLSWTEKEQNFSIINGSKRRTARGSGTIGEVNLLKQF 419  
  
QY 421 NDMKMMKQFTGGGK-KGKRNQNMNKLGMNLP 455  
DB 420 BETKMMKRFADIKDLKRGK-----LRLPF 445  
  
RESULT 8  
Q9CF65 PRELIMINARY; PRT; 518 AA.  
AC Q9CF65;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE Signal recognition particle protein Ffh.  
GN FFH OR L1616.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1360;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=IL1403;  
RX MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wincker P., Mauger S., Jailion O., Malarne K.,  
RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
lactis ssp. lactis IL1403.";  
RL Genome Res. 11:731-753(2001).  
DR EMBL; AE006391; AK05714.1; -.  
DR HSSP; O07347; IFFH.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR000897; SRP54.  
DR InterPro; IPR004125; SRP54\_SBP.  
DR InterPro; IPR004780; SRP\_sub.  
DR Pfam; PF00448; SRP54; 1.  
DR Pfam; PF02881; SRP54\_N; 1.  
DR Pfam; PF02978; SRP\_SBP; 1.  
DR ProDom; PD000819; SRP54; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMS; TIGR00959; 3a0501s01; 1.  
DR PROSITE; PS00300; SRP54; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 518 AA; 56845 MW; BF4539EF037601B4 CRC64;  
  
Query Match 57.8%; Score 1315; DB 16; Length 518;  
Best Local Similarity 55.4%; Pred. No. 1.2e-62;  
Matches 258; Conservative 88; Mismatches 102; Indels 18; Gaps 2;  
  
QY 1 MAFEGLSERLOATMOKMRGKGLTEADIKIMREVRLEALFEADVNFVKVKEFIKTVSERA 60  
DB 1 MAFENLTERLQNVFKNLGKKTTETDVTETKTEIRVALLEADVLPVVKFIKAIERA 60  
  
QY 61 LGSVMQSLTPGGQVQIKIVODELTKLGGENTSNMNSKPPTVYMMVGLQGAGKTTAGK 120  
DB 61 VGVEVSEALNPAQOVQIKIVDEELTAILGGGEAEELKSPKPTIIMVVGLOGAGKTTFAGK 120  
  
QY 121 LALLMRKYNKPKMLVAADYIPRAINOLOTVGQIDIPVYSEGQVKKPQIVTNALKHA 180  
DB 121 LAKKLEEQNARPLMTAADYIPRAIDQLKTLGOLBIPYVDGTAEEKPVNIVRNLKA 180  
  
QY 181 KEHLDFVIIDTAGRLHIDEALMNELEKVEIAKPNELMVDVMTGQDVNVAESFDDQ 240  
DB 181 QBERKDYVLIDTAGRLHIDEALMNELEKVEIAKPNELMVDVMTGQDVNVAESFDDQ 240  
  
QY 241 LDVGTVLTLDGDTGGAALSIRSVTKPIKVFVGMSEKLDGLFELPHPERMASRILMGD 300  
DB 241 LDITGVLTLDGDTGGAALSIRSVTKPIKVFVGMSEKLDGLFELPHPERMASRILMGD 300  
  
QY 301 VLSLIEKAQDVQDEKAKOLEKKMRESSFTLDDFLEQLDQVKNLGPLDDIMKMPGNMK 360  
DB 301 MTLTIEKAQANYDEEQSAKLAEKMAENRFYEDFVEQLDQVTNMGPMEDIMKMPGMSQM 360  
  
QY 361 KGLDLKLNSEKQIDHIKAIQSWTPAERNPDTLNVRKKRIAKGSGRSLQEVNRLMKQF 420  
DB 361 PGLDKVKVDFKRVARKAMVLSNTPAERHLEAELSPARRRRRRAAGSGNFIENKFIKQF 420  
  
QY 421 N-----DMKMMKQFTGGGKKGKRNQNMNKLGMNLP 454  
DB 421 NOSKEMWOGIMNGDMNAMQKMMGGAGG-----ONPNMPAGSGMP 460  
  
RESULT 9  
Q97QD2 PRELIMINARY; PRT; 523 AA.  
AC Q97QD2;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DE Signal recognition particle protein.  
GN SPI287.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1313;









GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 01:10:37 : Search time 1563 Seconds  
(without alignments)  
4714.617 Million cell updates/sec

Title: US-09-943-108A-2  
Perfect score: 2274  
Sequence: 1 MAPGLSERLQATMKQKRGK.....GKKGRNQNMKGNLPP 455

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlip  
-MODEL=frame+p2n.model -US09943108/runat\_21022003\_151631\_1824/app\_query.fasta\_1.647  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCTL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USRF=US09943108.ecgn\_1\_1\_2874/runat\_21022003\_151631\_1824 -NCPU=6 -ICPU=3  
-NO\_XLPAY -NO\_MAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRESH=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_man:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
c 1	1074	47.2	2169	17	BH770912	BH770912 LLMGtag64
c 2	757.5	33.3	875	17	AF075878	AF075878 AF075878
c 3	715.5	31.5	773	13	BI934118	BI934118 EST554007
c 4	631	27.7	656	14	BQ805434	BQ805434 WHE3566_H
c 5	630	27.7	711	12	BF940797	BF940797 7D99a06_x
6	616.5	27.1	2387	11	AK011928	AK011928 Mus muscu
7	594.5	26.1	688	17	BH394904	BH394904 AG-ND-138
8	562.5	24.7	605	10	BE432304	BE432304 EST398833
9	558	24.5	593	10	AW648731	AW648731 EST327101
10	555.5	24.4	581	9	AI977909	AI977909 496035806
11	534	23.5	677	13	BM413418	BM413418 EST587745
12	520.5	22.9	1200	11	AY109321	AY109321 Zea mays
13	513.5	22.6	533	10	AW930742	AW930742 EST356585
14	506.5	22.3	677	14	BQ996389	BQ996389 QGG12107
15	485	21.3	514	13	BM323434	BM323434 PIC1_19_C
16	479	21.1	1020	17	CNS0780K	AL437274 T3 end of
17	473	20.8	791	17	AQ327477	AQ327477 nbxb0041K
18	465.5	20.5	1032	11	AY106667	AY106667 Zea mays
19	463.5	20.4	483	10	BE596266	BE596266 P11_51_F0
20	451	19.8	777	17	BH402176	BH402176 AG-ND-122
21	448.5	19.7	490	13	BM325805	BM325805 PIC1_53_A
22	438.5	19.3	554	17	BH369841	BH369841 AG-ND-126
23	437.5	19.2	481	17	BH376496	BH376496 AG-ND-178
24	433.5	19.1	880	17	CNS077VW	AL433218 T7 end of
25	420	18.5	465	9	AI087737	AI087737 SWACAC23
26	419	18.4	476	10	AW033339	AW033339 EST276910
27	419	18.4	899	14	B0736950	B0736950 AGENCOURT
28	414.5	18.2	547	10	BE461473	BE461473 EST412892
29	412.5	18.1	688	13	BI933601	BI933601 EST553490
30	412.5	18.1	1108	9	AU091185	AU091185 AU091185
31	408	17.9	1059	14	BM806783	BM806783 AGENCOURT
32	405	17.8	1012	17	AF094895	AF094895 AF094895
33	400	17.6	949	17	CNS076TA	AL431828 T7 end of
34	398.5	17.5	814	17	AZ529841	AZ529841 ENT0084TR
35	392.5	17.3	889	17	AZ682189	AZ682189 ENT0084TR
36	389.5	17.1	773	13	BI454259	BI454259 603170687
37	388	17.1	761	9	AL655153	AL655153 AL655153
38	385.5	17.0	656	12	BF641333	BF641333 NF064F12I
39	380.5	16.7	694	13	BJ473333	BJ473333 BJ473333
40	378.5	16.6	874	12	BG679799	BG679799 602626480
41	377	16.6	707	17	AQ654845	AQ654845 Sheared D
42	375.5	16.5	753	13	BG921695	BG921695 602825514
43	375	16.5	382	10	AW649149	AW649149 EST327603
44	374.5	16.5	975	13	BG919869	BG919869 602823428
45	369.5	16.2	573	12	BG317280	BG317280 947027E11

ALIGNMENTS

RESULT 1  
BH770912/c  
LOCUS BH770912 2169 bp DNA linear GSS 01-MAY-2002  
DEFINITION LLMGtag64 MG1363 Random Sequence Tag Library Lactococcus lactis  
subsp. cremoris genomic, DNA sequence.  
ACCESSION BH770912.1 GI:20373869  
VERSION BH770912.1  
KEYWORDS GSS.  
SOURCE Lactococcus lactis subsp. cremoris.  
ORGANISM Lactococcus lactis subsp. cremoris  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Lactococcus.  
REFERENCE 1 (bases 1 to 2169)  
AUTHORS Bolotin,A., Ehrlich,S.D. and Sorokin,A.  
TITLE Studies of genomes of dairy bacteria Lactococcus lactis



JOURNAL  
COMMENT

Sci. Aliments, (2002) In press  
Contact: Sorokin A  
Genetique Microbienne  
INRA  
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France  
Tel: 33 1 34 65 25 16  
Fax: 33 1 34 65 25 21  
Email: sorokine@jouy.inra.fr  
best homologue in strain IL1403 is ffh (91%)  
Class: shotgun  
High quality sequence start: 30  
High quality sequence stop: 2141.  
Location/Qualifiers  
1. .2169  
/organism="Lactococcus lactis subsp. cremoris"  
/strain="MGI363"  
/db\_xref="taxon:1359"  
/clone\_lib="MGI363 Random Sequence Tag Library"  
/notes="Vector: pSGM2; Site\_1: SmaI; Library of  
chromosomal fragments of L.lactis strain MGI363 was  
prepared by partial AluI digestion or by sonication."  
BASE COUNT 642 a 429 c 380 g 718 t

## FEATURES

source

Alignment Scores:  
Pred. No.: 4.02e-108 Length: 2169  
Score: 1074.00 Matches: 218  
Percent Similarity: 73.86% Conservative: 73  
Best Local Similarity: 55.33% Mismatches: 82  
Query Match: 47.23% Indels: 21  
DB: 17 Gaps: 2  
US-09-943-108a-2 (1-455) x BH770912 (1-2169)

QY 76 ileLysileValGlnaspGluLeuThrLysLeuMetGlyGlyGluasnThrSerileasn 95  
Db 2168 ATCAAAATGTGTAGTGAAGAAATACAACTATCTTAGCGGTGGTGAACCCGAATACTC 2109  
QY 96 MetSerAsnLysProProThrValValMetMet-VaigLysLeuGlnGlyValagLysTh 115  
Db 2108 AATCTCTTAAATCCCAACAATATCATGATGGTGGTTCGACAGGGCAGGTAAAC 2049  
QY 115 rThrThrAlaGlyLysLeuAlaLeuLeuMetArgLysLysThrAsnLysPrometLe 135  
Db 2048 AACCTTCTGCTGTAATACTGCAAAAACTTAAAGAAACAATAATGCTGCCACTGAT 1989  
QY 135 uVal-AlaAlaAspIleThrArgProAlaAlaIle-AsnGlnLeuGlnThrValGlyLys 154  
Db 1988 GATTGGCAGCGAGCTTATPCGTCAGCAGCGATGGACCAATTAACCACTTGGAGAA 1929  
QY 155 GlnIleAspIleProValTySerGluGlyAspGlnValLysProGlnGlnIleValThr 174  
Db 1928 CAACCTGACATCTCTGTATGATGAAGAACGCTCTGAAAACCCGTTAATATTGTCGT 1869  
QY 175 AsnAlaLeuLysHisAlaLysGluGluHisLeuAspPheValIleleAspThrAlaGly 194  
Db 1868 AATGTCCTTTTAAAGCTCAAGAACGCTAAAGATTATGTGTGATTGATACGCGAGGT 1809  
QY 195 ArgLeuHisIleAspGluAlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLys 214  
Db 1808 CGCTGTGAATGTGATACATATGATGATGATGATGATGATGATGATGATGATGATGAT 1749  
QY 215 ProAsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAla 234  
Db 1748 CCAACAGAAATTTACTCGTGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1889  
QY 235 GluSerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAsp 254  
Db 1688 AAAAATTTTGTGATGAAAAATAGATATACCGGTGTATCATCAACCACTCGATGGGAC 1629  
QY 255 ThrArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheVal 274  
Db 1628 ACACGTGTGTGGCTTTATCAATTCGTGAAATTAACCGGAAACCTTATAATTTACA 1569

QY 275 GlyMetSerGluLysLeuaspGlyLeuGluLeuPheHisProGluArgMetAlaSerArg 294  
Db 1568 GGGACTGCTGAAAAATTAACCGACCTTGAAGCTTTCTACCTGATCGTATGAGCTCAGA 1509  
QY 295 IleLeuGlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnaspValaspGln 314  
Db 1508 ATTCTGTGTATGGGTGATATGCTCACTGCTGATTGAGAAGCGCAAGCTAATTATGATGAG 1449  
QY 315 GluLysAlaLysaspLeuGluLysLysMetArgGluSerSerPheThrLeuAspPhe 334  
Db 1448 GAACAATCTGCAAACTCGCTGAAAAATGCTGATAACCCGTTTGTATTATGAGGATTC 1389  
QY 335 LeuGluGlnLeuaspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIle 354  
Db 1388 GTTGAGCACTAGATCAAGTTACAAATATGGGACCAATGGAAGATATTATGAATGATT 1329  
QY 355 ProGlyMetAsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleasp 374  
Db 1328 CCAGGAATGTCACAAATGCTGCTGCTGATAAAGTTAAAGTTGACCTTAAGATGTTGCT 1269  
QY 375 HisIleLysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeu 394  
Db 1268 AGAAACGACGAATGTTGTCATGACCCGCTGACCTGACCTGATTTAGAACGACAACTA 1209  
QY 395 AsnValSerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsn 414  
Db 1208 TCGCCAGCTGCTGCTGCTGCAATCGCCGAGGTCTGGGAATTCATTTATTGAAGTAAT 1149  
QY 415 ArgLeuMetLysGlnPheAsn----- 422  
Db 1148 AATTCATTAGCAATTTAATCAATCAACAAAGAAATGATCAAGGAATCATGAATGCTGAT 1089  
QY 423 MetLysLysMetLysGlnPheThrGlyGlyGlyLysLysLysLysLysLysLys 442  
Db 1088 ATGAATGCGATGATGCAAAAAATGATGGGTGGAGCTGGTGGG----- 1047  
QY 443 GlnMetGlnAsnMetLysGlyMetAsnLeuPro 454  
Db 1046 CAAATGCCAATATGCCACCGGTTCTGGAATGCCA 1011

## RESULT 2

AF075878/c

LOCUS

AF075878

DEFINITION

typhimurium genomic clone 245-T3, DNA sequence.

ACCESSION

AF075878

VERSION

AF075878.1

KEYWORDS

GSS.

SOURCE

Salmonella typhimurium.

ORGANISM

Salmonella typhimurium.

REFERENCE

1 (bases 1 to 875)

AUTHORS

Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.

TITLE

Sample sequencing of a Salmonella typhimurium LT2 lambda library:

comparison to the Escherichia coli K12 genome

JOURNAL

FEMS Microbiol. Lett. 173 (2), 411-423 (1999)

MEDLINE

99243757

COMMENT

Contact: McClelland M

Molecular Biology

Sidney Kimmel Cancer

3099 Science Park Road, San Diego, CA 92121, USA

Email: mclelland@lifsci.sdsu.edu

Class: shotgun.

Location/Qualifiers

1. .875

/organism="Salmonella typhimurium"

/strain="LT2"

/db\_xref="taxon:602"

/clone="245-T3"

/note="Vector: Lambda DASH II; sequenced using Li-Cor

BASE COUNT	178 a	266 c	217 g	214 t
ORIGIN	sequencer"			
Alignment Scores:	1.51e-73	Length: 875		
Pred. No.:	757.50	Matches: 152		
Score:	73.08%	Conservative: 57		
Percent Similarity:	53.15%	Mismatches: 76		
Best Local Similarity:	33.31%	Indels: 2		
Query Match:	17	Gaps: 1		
DB:				
US-09-943-108a-2 (1-455) x AF075878 (1-875)				
QY	87	MetClyGlyCluAsnThrSerIleAsnMetSerAsnLysProProThrValValMetMet	106	
Db	867	ATGGGGAAGAGAACACACGCTGAATTTTAC-TGCAGCGCCAGCGGTAGTATTGATG	809	
QY	107	ValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArg	126	
Db	808	CGGGTCTGACGGGCGGGTAAACACACGCGTGGTAAATTCCTTGGCG	749	
QY	127	LysLysThrAsnLysLysProMetLeuValAlaAlaAspIleTyArgProAlaAlaIle	146	
Db	748	GAGAAACACAGAGAAAGTGTGCTCTCTGCGACGTATATGCGCGGGCGATC	689	
QY	147	AsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTySerGluGlyAspGln	166	
Db	688	AACAGCTCGAAGCTGGCTGACGAGGTGGCGTGAATTCCTCCGCTGTGATGTCGC	629	
QY	167	ValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAsp	186	
Db	628	CAGAAACCGGTTGATATCGTCAACGCGCGCTGAAAGAGCGAAGCTCAATTCACGAC	569	
QY	187	PheValIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeu	206	
Db	568	GTCTGTCTGTGTATACCGCGGTCTGTCATGTTGACGAGCGATGATGGACGAATC	509	
QY	207	LysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSerMetThr	226	
Db	508	AACAGCTCCACGCTTATCAACCCAGTAGAAGACGCTTTGTCGTGATGGATGAC	449	
QY	227	GlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyVal	246	
Db	448	GTCAGATGCGCGGATACCGCAAGCCTTTAACGAGAGCGCTGCCCTTAACGCGCG	389	
QY	247	ThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArgSerVal	266	
Db	388	GTGCTGACCAAGTTGATGATGATGCGGTGGCGGTGGCGGCTCTCTATTGTCATATC	329	
QY	267	ThrGluLysProIleLysPheValGlyMetSerGluLysLysLeuAspGlyLeuGluLeuPhe	286	
Db	328	ACCGGACGCGATTAATTCCTCGGTGTCGGGAGAAACGACGCGCTGGAGCCATTC	269	
QY	287	HisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIleGlu	306	
Db	268	CATCCGGATCGTATCGCTCCGCTATCTCGGATGGCGGACCTACTGTCTCTATCGAA	209	
QY	307	LysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGluLysLysMetArgGlu	326	
Db	208	GATATGAAAGCAAGTTGACCGGCGACAGGCTGAGAGCTGGCGACCAACTGAAGAA	149	
QY	327	---SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGly	345	
Db	148	GGCGAGGTTTCGACTGAACGACTTCTCTGAAACAGCTCAACAGATGAAGAAACATGGG	89	
QY	346	ProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLys	365	
Db	88	GGTATGGCCAGCTGATGGGCAAAATACCGGCGATGGCGAGATTCGGCAACAGTTAA	29	
QY	366	LeuAsnMetSerGluLys	371	
Db	28	TCGACAGATGATCACAG	11	

RESULT 3

BI934118

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

tomato.

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 773)

van der Hoeven,R.S., Bezzerides,J.L., Karanymcheva,S.A., Tsai,J., Uterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

Generation of ESTs from tomato flower tissue, anthesis (2001)

Unpublished (2001)

Contact: CGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics Institute

Seq primer: T3.

Location/Qualifiers

1..773

/organism="Lycopersicon esculentum"

/cultivar="TA96"

/db\_xref="taxon:4081"

/clone="cTOD18J16"

/tissue="tomato flower, anthesis"

/dev\_stage="flower"

/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; supplier: Corneil University; sequencing: The Institute for Genomic Research; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA96). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

BASE COUNT

227 a

129 c

210 g

207 t

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

5.83e-69

715.50

75.19%

51.94%

31.46%

13

Length: 773

Matches: 134

Conservative: 60

Mismatches: 63

Indels: 1

Gaps: 1

US-09-943-108a-2 (1-455) x BI934118 (1-773)

QY

39

LeuPheGluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGlu

58

Db

1

CTTTTAGAAGCTGATGTAGTCTCCAGTTGTCAGAAAGGTTGTTTCAGTCTGTAGTGAA

60

QY

59

ArgAlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIle

78

Db

61

GAGCGGTTGGCAGCTGGTTGATTCGAGAGTGTAAACACAGATCAGACTAGTTAAAGCT

120

QY

79

ValGlnAspGluLeuThrLysLysMetGlyGlyGluAsnThrSerIleAsnMetSerAsn

98

Db

121

GTACGTGACAGCTTGTGAACATGATGGTGGAGAGGTTCTTGAACTGGTTTTTGTCTAA

180

QY

99

LysProProThrValValMetMetValGlyLeuGlnGlnValAlaGlyLysThrThrAla

118

Db

181

TCATAGCCCAACCGTAACTACTATTGGCCGGTCTACAGGTTGTGGGAGACACACTGTTAGT

240

```

QY 119 GlyLysLeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAla 138
Db 241 GCAGAGTGTATATCTA---AGAGACAGGGTANAGATTGCCATGCTGATTCTCGA 297

QY 139 AspLeuTyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLeuAspLeu 158
Db 298 GACGTGTACAGACCTGCTGATTATGACCAACCTGTTATTTGGTAAACAGGTTGATGTA 357

QY 159 ProValTyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLys 178
Db 358 CTGTTTATGACAGCAGACAGATGTAAACCTCGCAATATAGCCGACCAAGATTAACA 417

QY 179 HisAlaLysGluGluHisLeuAspPheValLeuLeuAspThrAlaGlyArgLeuHisLeu 198
Db 418 GAGGCCAAAAGAAAGATGTAGATGTAGTCATATGATGACAGCTGGAAAGACTTCAGATA 477

QY 199 AspGluAlaLeuMetAsnGluLeuLysGluValLysGluLeuLeuAlaLysProAsnGluLeu 218
Db 478 GATAAACTATGATGGATGAATTAAGAACGCTGAACGGGTACTGACACCCACAGAGGTT 537

QY 219 MetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAsp 238
Db 538 TTAGTTGTGTGATGATGATGCTGCGCAGAGCTGCACGCTTGTGTCACACATCAAT 597

QY 239 AspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGly 258
Db 598 CTCGAATTTGGATTACTGTGCTCATCTTACGAGAGCTAGATGAGATCTTAGGGGTGA 657

QY 259 AlaAlaLeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGlu 278
Db 658 GCAGCTTTAGTGTACAGGAGGATCAGGAAGCCAAATCAAGCTTCGTAGGAAGGGGTGAA 717

QY 279 LysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgLeuLeu 296
Db 718 CGATGGAGGACCTTGACCTTCTATCTGACCGCATGGCTGCAGCTATTTTA 771

RESULT 4
BQ805434
LOCUS
DEFINITION
WHE3566_H12_P2425 Wheat developing grains cDNA library Triticum
aestivum cDNA clone WHE3566_H12_P24, mRNA sequence.
ACCESSION
BQ805434
VERSION
BQ805434.1 GI:22029643
SOURCE
EST.
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 656)
Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin
,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J.,
Wilson,C. and Woo,J.
The structure and function of the expressed portion of the wheat
genomes - Developing grains cDNA library
Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequences have been trimmed
Quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES
source
Location/Qualifiers
1..656
/organism="Triticum aestivum"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3566_H12_P24"
/clone_lib="Wheat developing grains cDNA library"

```

```

/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/notes: Vector: Lambda ZAP II, excised phagemid; Site_1:
ECORI; Plants were grown under six following different
environmental regimes in greenhouse, Environment 1)
240C/170C day/night, well-watered, with post-anthesis
fertilizer, Environment 2) 240C/170C day/night,
well-watered, without post-anthesis fertilizer,
Environment 3) 370C/170C day/night, well-watered, with
post-anthesis fertilizer, Environment 4) 370C/170C
day/night, well-watered, without post-anthesis fertilizer,
Environment 5) 370C/170C day/night plus drought, with
post-anthesis fertilizer, Environment 6) 370C/170C
day/night plus drought, without post-anthesis fertilizer,
developing wheat grains from the following were excised
and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28,
32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24
, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16,
20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16
, 20, 24, 28, 30 DPA and total RNA was prepared by S.
Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA
library was made using poly (A) RNA, and the cDNA clones
were in vivo excised to give pBluescript SK(-) phagemids
in the TJ Close lab (Chih, close, Fenton) at the
University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (others)."
BASE COUNT 200 a 115 c 179 g 162 t
ORIGIN
Alignment Scores:
Pred. No.: 11e-59 Length: 656
Score: 631.00 Matches: 116
Percent Similarity: 75.23% Conservative: 48
Best Local Similarity: 53.21% Mismatches: 54
Query Match: 27.75% Indels: 0
DB: 14 Gaps: 0
US-09-943-108A-2 (1-455) x BQ805434 (1-656)
QY 141 TyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLeuAspLeuProVal 160
Db 3 TACAGCCTGCTGCCATGTATCACTACTGTACTGGTGAACAGGTTGGTGCACGTT 62
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180
Db 63 TACTCAGAAGGACCGCGGCCAACCTGCAGAAATPAACCAAGATCCGTGGAAGAGCG 122
QY 181 LysGluGluHisLeuAspPheValLeuLeuAspThrAlaGlyArgLeuHisLeuAspGlu 200
Db 123 AAAGAATAATATGTATGCGATGCGATGATGATGATGATGATGATGATGATGATGAT 182
QY 201 AlaLeuMetAsnGlnLeuLysGlnValLysGlnLeuAlaLysProAsnGlnLeuMetLeu 220
Db 183 ACAATGATGTTGATTAATGAAGAGATGAAGAGGACGATTAATCTACAGATTTGGTT 242
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 243 GTTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 303 ATTGTATATCCGTGCTATATTGATTAATGATGATGATGATGATGATGATGATGATGAT 362
QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGlnLysLeu 280
Db 363 CTATGCTGAAGAGGCTGCTGGAAGCCCATCAAGTTTGTGGCGAGAGAGCGAATG 422
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgLeuLeuGlyMetGlyAsp 300

```

```

Db 423 GAGGACCTGAGCTTTCATCCGACCGCATGGCACACCTGTTTGGGAATGGAGAT 482
    ... ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValaspGlnGluLysAlaLysAspLeu 320
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 483 GTCCPTTCATTTTGTGAAACACACAGAACCTGCTCCCAAGAGATACCATGGAATG 542
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 321 GluLysMetArgGlnSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 543 CAGAGAGATCATGAGTGGGAATTTGACTTCATGCTTTTAAAGCAGACACAAAT 602
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 341 ValLysAsnLeuGlyProLeuAspPheMetLysMetIleProGlyMetAsn 358
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 603 GTTGGAAATGGGATCCAGAGCCGTGTATCGGAATGATCCAGGCATGAAC 656
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 5
BF940797/c 711 bp mRNA linear EST 22-JAN-2001
LOCUS BF940797
DEFINITION 7d99a06.x1 Lupski_dorsal_root_ganglion Homo sapiens cDNA clone
IMAGE:3280811.3' similar to SW:SR54_BACSU P37105 SIGNAL RECOGNITION
PARTICLE PROTEIN ; mRNA sequence.
ACCESSION BF940797
VERSION BF940797.1 GI:12358117
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 711)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccgaps-remail.nih.gov
Tissue Procurement: Dr. James Lupski
cDNA Library Preparation: Lupski Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 433.
FEATURES
source
Location/Qualifiers
1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3280811"
/clone_lib="Lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/lab_host="DH10B"
/dev_stage="adult, 36 yr"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGATGCTTAGTCGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 193 a 145 c 88 g 283 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 1.57e-59 Length: 711
Score: 630.00 Matches: 134
Percent Similarity: 77.45% Conservative: 48
Best Local Similarity: 57.02% Mismatches: 45
Query Match: 27.70% Indels: 8

```

```

Db: 12 Gaps: 3
US-09-943-108a-2 (1-455) x BF940797 (1-711)
Qy 202 LeuMetAsnGluLeuLysGluValLysGluLysProAsnGluLeuMetLeuVal 221
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 708 TTAATGCCAGAGTAAAGATGTTAAAGAGTTTCAATCTAGTGAATATTATTAGTA 649
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 222 ValAspSerMetThrGlyGlnAspAlaValAsnValLalGluSerPheAspAspGlnLeu 241
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 648 GTTGATGCTATGACAGCTCCAGATGCTGTTAATGTAGCAGAGACATTTAATACATCA 589
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 242 AspValThr-GlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaLale 261
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 588 GATCTTTCAAGNGAATAATATATACAAAGTTAGATGGTGATACAGAGGTGGTCTCT 529
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 261 uSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAs 281
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 528 TTCAATAAGAGACATTAATGCGCAAGCAATTAATAATTTGTTGGTGGTCAACAGATGAG 469
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 281 pGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetClyAspVa 301
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 468 TGATATTGAAGTATTCATCCAGATAGATGCTTCAAGATATATTAGGAATGGGAGATGT 409
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 301 lLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGl 321
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 408 CTTTCTTTAATAGAAAAAGCTCAGCAAGCTATTGACCAGATGACCAAGTAAGTTAAG 349
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 321 uLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGlnVa 341
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 348 TGAATAAATGTTAAATCAAGAATTTAACTTTGATGACTACTTATCACTGATGATCAAT 289
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 341 lLysAsnLeuGlyProLeuAspPheMetLysMetIleProGlyMetAsn-Lys---M 360
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 288 GAAAAAGCTTGGACCTATAATAATTTAGATAGATGATTCACAGGTGTAAACAAAGAAC 229
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 360 etLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleI 380
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 228 TTGAGGCAATTGAT---TTTCTCAGAGGAGAAAAACAATAATGGCTACAGTTAAGCAATCA 172
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 380 leGlnSerMetThrProAlaGluArgAsnProAspThrLeu-----AsnValSera 398
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 171 TACAATCGATGACAGCTAAAGAAAGAAAAACAACCTTCTTTAGTAATAGGAATGGTTCAA 112
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 398 rGlyLysArgIleAlaLysGlySerGlyArgSerLeuGlnGlu-ValAsnArgLeuMet 417
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 111 GAAAGAGAAGATAGCTAAAAGTTCTGGTACCAACAGTACAAAGAGNTAAATAAAGTTCTT 52
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 418 LysGln-PheAsnAspMetLysLysMetMetLysGln 429
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 51 AAAGGCTATGAAATGATGAAAAAGCAATGAGCAA 15
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 6
AK011928 2387 bp mRNA linear HCT 19-JAN-2002
LOCUS AK011928
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610209C12:signal recognition particle 54
kda, full insert sequence.
ACCESSION AK011928
VERSION AK011928.1 GI:12848353
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA,
clone:lib:RIKEN full-length enriched mouse cDNA library
clone:2610209C12.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

```



```

QY 161 TyrSerGluGlyAspGlnValTyrProGlnGlnLeuValThrAsnAlaLeuLysHisala 180
Db 715 TATGCACCTACTAGAAATGGATGCTGTCATCATCTCTCTGAAGAGTGGAGAAATTC 774

QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 775 AAAAAATGAATTTTGAATATATTTGTTGATCAAGTGTGTCACAAACAAAGAAC 834

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 835 TCTTTATTTGAAGAAATGCTTCAAGTTCTTAAGCTATACACCTGATACATTTTAT 894

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 895 GTGATGATGATCATCATCGACAGGCTTGTGAGGCCCGAGGCGCTTTAAGACAA 954

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAla 260
Db 955 GTAGATGTAGTTCAGTAATAGTGACAAACTCGACGCTCATGCAAGAGGGCGGTCT 1014

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 1015 CTATGTCAGTGTGTCACCAAAAGTCCATTTATTTCTGTCACAGGGAACATATA 1074

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 1075 GATGATTTTGAACCTTTCAAAACACAAACCTTTCATCAGCAACTCCTTGGATGGTGT 1134

QY 301 ValLeuSerLeuIleGluLysAlaGlnGln---AspValAspGlnGlnLysAlaLysAsp 319
Db 1135 ATTGAAGGACGTGATTGATAAAGTCAATGAATTAAGTTGATGATGATAATGAGCA----- 1188

QY 320 LeuGluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAsp 339
Db 1189 CTTATAGAGAGTTTGACGACGGTCAAGTTACATTCGACGACATGATGACAGATTTTCCAG 1248

QY 340 GlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsn--- 358
Db 1249 AATATTGAAATGGGCCCATTCAGTCAGATATTGGGGATGATCTCTGGCTTGGCACA 1308

QY 359 -----LysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIle 376
Db 1309 GATTTTATGACAAAGGAATGACAGAGTCAATGGCA-----AGGCTGAAGAAATG 1362

QY 377 LysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnVal 396
Db 1363 ATGCAATCATGGACATATGACGATCAAGAACTGACAGTACAGATGTCGCAAGGTT 1422

QY 397 -----SerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGln 411
Db 1423 TTCAGTAAGCAACACGAGGAGATCCAAAGAGTTGCCCGGGATCAGGTGTGTCAACAAGA 1482

QY 412 GluValAsnArgLeuMetLysGlnPheAsnAspMetLysLysMetLysGln----- 429
Db 1483 GATGTTCAAGAACCTCTGACCCAGTATACCAAGTTTGCACAGATGTCGCAAAAGATGGGA 1542

QY 430 -----PheThrGlyGlyGlyLysGlyLysLysGlyLysArgAsnGlnMet 444
Db 1543 GGTATCAAGGACTTTTCAAGCGGTGATATGCTATAGATGTGAGTCAAGTCAAGATG 1602

QY 445 -----GlnAsnMetLysGlyMetAsn 452
Db 1603 GCAAAATTAACCAACAAATGGCCAAATGATGGAC 1638

RESULT 7
BH394904
LOCUS BH394904
DEFINITION AG-ND-138B11.TF ND-TAM Anopheles gambiae genomic clone AG-ND-138B11
, DNA sequence.
ACCESSION BH394904
VERSION BH394904.1 GI:17341045
KEYWORDS GSS.
SOURCE African malaria mosquito.

```

```

ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE 1 (bases 1 to 688)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-138B11.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PGST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seg primer: M13 For
Class: BAC ends.
Location/Qualifiers
source 1..688
/organism="Anopheles gambiae"
/strain="PGST"
/db_xref="taxon:7165"
/clone="AG-ND-138B11"
/clone_lib="ND-TAM"
/vector="pECBAC1; Site_1: HindIII"
BASE COUNT 223 a 118 c 154 g 193 t
ORIGIN
Alignment Scores:
Pred. No.: 1,29e-55 Length: 688
Score: 594.50 Matches: 129
Percent Similarity: 73.39% Conservative: 42
Best Local Similarity: 55.36% Mismatches: 55
Query Match: 26.14% Indels: 7
DB: 17 Gaps: 1
US-09-943-108A-2 (1-455) x BH394904 (1-688)
Qy 60 AlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleVal 79
Db 2 GCTTTAGGACAAATGTTTAAACATCCATTACCCGGGACAAATGATGACAAAAATGTT 61
Qy 80 GlnAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLys 99
Db 62 CAGCATGATGTAGTGGATTTAATGGGAGGAACCAATGAAGGCTTAATCTTCCGAAAAA 121
Qy 100 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGly 119
Db 122 ---CCACAAATTAATACGTATGAGGTTTACAGGTTCTGGTAAACTCATCTCCCGGG 178
Qy 120 LysLeuAlaLeuLeuMetArgLysLysTyrAsnLysLysPrometLeuValAlaLaasp 139
Db 179 AAATTAGCAAACTTCTCTCAAGAGAAAGAGACAAATAATCGTTATGTTGATGATCTGAC 238
Qy 140 IleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIlePro 159
Db 239 GTTATATCTCTCTGCTGGGATGATGACACTAAAAGTATTAGGTAGTCAGACAGGGTCTCT 298
Qy 160 ValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHis 179
Db 299 GTATATCTAGTAAGAGGTAACATAATCGGTTTCAGATTCTCAGATGCAATGAGATTT 358
Qy 180 AlaLysGluGluHisLeuAspPheValIleIleAspThrAla-GlyArgLeuHisIleAs 199

```

```

|||||:  ||| :|||:|||||:|||||:|||||:|||||:|||||
Db 359 GCAAAGCAAAATAAGCACGATGTTATCATATAGATACGCGGCGTTTGGCTATCGA 418

QY 199 pGIuAlaLeuMetAsnGIuLeuLysGIuValLysGIuIleAlaLysProAsnGlu-Ilex 219
|||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 419 TGAAGAGATGATGAACGAAATCTTAACGTTCACCTGCTGTAACACCTACAGAACTCT 478

QY 219 etLeuValValAspSerMetThrGIyGlnAspAlaValAsnValAlaGluSerPheAspA 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 479 TTTTCGTAGTTGACTCTATGACTGGGAGGATGCTGTGATCCAGCAAAAGCTTTTAACG 538

QY 239 spGlnLeuAsp-ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArg-GlyEl 258
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
Db 539 ATGFCCTTAATATACGGGTGATGTTACTTAATAGATGGTGATACGCGGAGGTGG 598

QY 258 yAlaAlaLeuSerIleArgSer-ValThrGlnLysProIleLysPheVal-GlyMetSer 277
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
Db 599 GGGCGGCATAACATCCGTTCCGGTAGTAGAAAAACCAATTAAGTTAATTTCTACCGGT 658

QY 278 GlnLysLeuAspGlyLeuGluLeuPhe 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 659 GAGAAATGATGCTCTGGGATATTTT 685

RESULT 8
BE432304
LOCUS BE432304 605 bp mRNA linear EST 18-MAY-2001
DEFINITION EST398833 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
ACCESSION BE432304
VERSION BE432304.1 GI:9430147
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 605)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.

FEATURES
source
location/Qualifiers
1..605
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG7M2"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="Pericarp"
/lab_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCudapt; Site.1: EcoRI;
Site.2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 190 a 97 c 159 g 159 t
ORIGIN

Alignment Scores:
Pred. No.: 3,86e-52 Length: 605
Score: 562.50 Matches: 105
Percent Similarity: 75.25% Conservative: 47
Best Local Similarity: 51.98% Mismatches: 49

```

```

Query Match: 24.74% Indels: 1
Ds: 10 Gaps: 1

US-09-943-108A-2 (1-455) x BE432304 (1-605)

QY 65 ValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGluLeuThr 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2 TTGATTTCGAGGTGTAAACCAGATCAGCAACTAGTTAAGACTGTACGTGACGAGCTGTG 61

QY 85 LysLeuMetGlyGlyGlnAsnThrSerIleAsnMetSerAsnLysProProThrValVal 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 62 AAATCTGATGGGTGGAGAGGTTTCTGAACTGGTGTGTTTCTAAATCTAAGCCACCAGTA 121

QY 105 MetMetValGlnGlnGlnValAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeu 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 122 CTATTGCCGGCTACAGAGGTGTGGGAAGACAACCTGTAGTGCAAGTTAGCTTATAT 181

QY 125 MetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIleTyrrArgProAla 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 182 CTA--AAGAAGCAGGGTAAGAGTTGCATGCTGCTGGAGAGCTGTACAGACCTGCT 238

QY 145 AlaIleAsnGlnLeuGlnThrValGlyGlnGlnIleAspIleProValTyrrSerGluGly 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 239 GCTATTGACCAACTGTTATTGTTGGTAAACAGGTTGATGATACCTGTTTATGTCAGACGA 298

QY 165 AspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHis 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 299 ACAGATTAACACCTGCAGAAATAGCCCGACAGAGATTACAAGAGCCCAAAAAGAAAT 358

QY 185 LeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsn 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 359 GTAGATGTAGTCATATGATACAGCTGGAAGACTTCAGATAGATAAAACTATGATGATG 418

QY 205 GluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSer 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 419 GAATTAAGACGCTGAACGGGTACTGAACCCACAGAGGTTTACTTTGTTGGATGCA 478

QY 225 MetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThr 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 479 ATGACTGGCCAAAGACTCGACGCTTTGTCACACATTCATCTCGAAATGGGAATFACT 538

QY 245 GlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArg 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 539 GGTCCCATCTTGAGAGAGCTAGATGGAGATCTTAGGGGTGGAGAGCTTTAAGTGTCAA 598

QY 265 SerVal 266
|||
Db 599 GAGGTA 604

RESULT 9
AW648731
LOCUS AW648731 593 bp mRNA linear EST 18-MAY-2001
DEFINITION EST327101 tomato germinating seedlings, TAMU Lycopersicon
esculentum cDNA clone cLEI5G22 5', mRNA sequence.
ACCESSION AW648731
VERSION AW648731.1 GI:7409885
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 593)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University

```





```

:::|||||
Db 361 GTACCAGTTTACTCAGAGGAAGTGAAGCAAAACCTTCACAATAAGCCAAAAGGGTTG 420
QY 178 LysHisAlaLysGluLHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHis 197
Db 421 AAGGAGCAAAAGTCAATAGCGTGTATTAATTATAGTGACACGGCTGGAAGACTGCAG 480
QY 198 IleAspGluAlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlu 217
Db 481 GTAGATAAACAATGATGATGAGTGAAGAGAGTAAAAAGACAGTGAATCCCTACAGAA 540
QY 218 IleMetLeuValLysSerMetThrGlycInAspAla 230
Db 541 GTTCTTCTGCGTTCATGCCATGACTGGCCCAAGAAGCT 579

RESULT 11
LOCUS BM4113418 677 bp mRNA linear EST 22-JAN-2002
DEFINITION EST587745 tomato breaker fruit Lycopersicon esculentum cDNA clone
cLEGG63C14 5' end. mRNA sequence.
ACCESSION BM4113418
VERSION BM4113418.1 GI:18265048
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 677)
AUTHORS Alcalá,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
,J., Bougria,O., Kirkness,E., Utterback,T., Van Aken,S., Rønning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
CONTACT: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
Location/Qualifiers
source 1..677
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEGG63C14"
/clone_lib="tomato breaker fruit"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOIR"
/note="Vector: pBluescriptSKmCudapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
Sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT 222 a 119 c 165 g 171 t
ORIGIN

Alignment Scores:
Pred. No.: 6,42e-49 Length: 677
Score: 534.00 Matches: 101
Percent Similarity: 75.38% Conservatve: 46
Best Local Similarity: 51.79% Mismatches: 47
Query Match: 23,48% Indels: 1
DB: 13 Gaps: 0

US-09-943-108a-2 (1-455) x BM4113418 (1-677)

```

```

QY 152 ValGlyLysGlnIleAspIleProValTySerGluGlyAspGlnValLysProGlnGln 171
Db 92 GTGGGGGAACAGGTGTGATCTACCTGTTTATGCAGCAGGAACAGATGTAACACCTGCAGAA 151
QY 172 IleValThrAsnAlaLeuLysHisAlaLysGluLHisLeuAspPheValIleIleAsp 191
Db 152 ATAGCCGACAGAGATTACAAGAGCCCAAAAGAAGATGTAGATGTAGTCATTAATGAT 211
QY 192 ThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLysGluValLysGlu 211
Db 212 ACAGCTGGAGAGCTCAGATAGATAAACTATGATGATGAATTAAGAGAGCTGAACGG 271
QY 212 IleAlaLysProAsnGluIleMetLeuValLysSerMetThrGlycInAspAlaVal 231
Db 272 GTACTGAACCCACAGAGAGTTTACTTGTGTGGATGCAATGACTGGCCCAAGAAGCTGCA 331
QY 232 AsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeu 251
Db 332 GCTTTGGTCACAACTCAATCTCAAAATTGGAATTACTGGTGGCCATCTTGACGAAGCTA 391
QY 252 AspGlyAspThrArgGlyAlaAlaLeuSerIleArgSerValThrGluLysProIle 271
Db 392 GATGAGAGATTCTAGGGGTGGAGCAGCTTTAAGTGTCAAGGAGGTATCAGGAAGCCANTC 451
QY 272 LysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMet 291
Db 452 AAGCTCGTAAGAAGGGGTGAAGGTATGAGGAGACTTGAACTTTCTATCTGACCGCATG 511
QY 292 AlaSerArgIleLeuGly-MetGlyAspValLeuSerLeuIleGluLysAlaGlnGlnAs 311
Db 512 GCTGGAGCTATTATTGAAGAATGGGAGATGTTCTATGCTTTGTGAGAAAGCCCAAGAA 571
QY 311 pValAspGlnGluLysAlaLysAspLeuGluLysLysMetArgGluSerSerPheThrLe 331
Db 572 TATGCTCAAGAGAGTCTGAGAGAAATTCAGACAGACATCATGCTGCAAAATTGATTT 631
QY 331 uAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGly 345
Db 632 CAATGACTTCTGAGCAAACTCCGGCAGTTGCTAGGATGGGT 674

RESULT 12
LOCUS AY109321 1200 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays P00070230 mRNA sequence.
ACCESSION AY109321
VERSION AY109321.1 GI:21212876
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1200)
AUTHORS Coe,E.C.
Direct Submission
TITLE Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
LOCATION/Qualifiers
source 1..1200
/organism="Zea mays"
/db_xref="MaizeDB:633818"
/db_xref="taxon:4577"
/clone="PC0070230"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public

```





below Phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with PolyTMix or T7 sequencing primer, are presented as the reverse complement. Seq primer: JEN REV High quality sequence stop: 454 POLYA-No.

FEATURES source

Location/Qualifiers  
1..514  
/organism="Sorghum bicolor"  
/cultivar="BTX623"  
/db\_xref="taxon:4558"  
/clone\_lib="Pathogen-infected compatible 1 (PIC1)"  
/tissue\_type="Leaves"  
/dev\_stage="4-week-old seedlings infected with Colletotrichum graminicola"

/note="Vector: pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; Four-week-old sorghum seedlings were sprayed with spore suspension prepared from 3-week-old FRM421, a sorghum isolate of the anthracnose pathogen Colletotrichum graminicola. Inoculated plants were kept in a 25 C dark growth chamber with 100% relative humidity for 24 hr, followed by 12/12 hr of light/dark cycle at 25 C with 90% relative humidity for another 24 hr. All leaves were harvested and quick frozen with liquid nitrogen and stored in a -80 C freezer. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 159 a 91 c 137 g 127 t  
ORIGIN

Alignment Scores:

Pred. No.: 1.22e-43 Length: 514  
Score: 485.00 Matches: 90  
Percent Similarity: 77.25% Conservative: 39  
Best Local Similarity: 53.89% Mismatches: 38  
Query Match: 21.33% Indels: 0  
DB: 13 Gaps: 0

US-09-943-108A-2 (1-455) x BM323434 (1-514)

QY 194 GlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLysGluValLysGluIleAla 213  
Db 10 GGAAGACTGCAGATTGATAAATCAATGATGATGAATGAAGAAGTGAAGAGCTGTT 69  
QY 214 LysProAsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnVal 233  
Db 70 RAATCTACAGAAGTTCGTGTCGTGATGCCATGACTGSCCAAGAAGCTGCACACTA 129  
QY 234 AlaGluSerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGly 253  
Db 130 GTCACACCTTCAATATTGAGATTGTTGATCTATCTGCTGCTATATTGCTAAATTGGATGTT 189  
QY 254 AspThrArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPhe 273  
Db 190 GACTCCAGGGCGGAGCCGACACTAGTGTAAAGAGCTCTCTGGGAAGCCCATCAAGTTT 249  
QY 274 ValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSer 293  
Db 250 GTTGGCGCTGGGAACGATTGGAGACCTTGAGCTTTCTACCCCTGATGCTGCATGGCACAG 309  
QY 294 ArgIleLeuGlyMetGlyAspValLeuSerLeuIleGlnLysAlaGlnAspValAsp 313  
Db 310 CGAGTATTGGGAATGGAGATGTACTCTCATTTGTTGAAAGACACAGAAGATTATGCGG 369  
QY 314 GlnGluLysAlaLysLeuLysLysMetArgGluSerSerPheThrLeuAspAsp 333  
Db 370 CANGAGAGGCTGTAGATTACAGAAAAGATCAGTACGTAATAATTCGACTTCAACGAC 429  
QY 334 PheLeuGluGlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMet 353

Db 430 TTCTTAACAATCTCAAAATTTCCGAAATGGGTTCATGAGCCGCAATTATTGGAATG 489  
QY 354 IleProGlyMetAsnLysMet 360  
Db 490 ATGCCAGGCATGAACAAGATA 510

Search completed: February 25, 2003, 03:09:37  
Job time : 1571 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:13:23 ; Search time 40 Seconds  
(without alignments)  
1515.726 Million cell updates/sec

Title: US-09-943-108a-2

Perfect score: 2274

Sequence: 1 MAPEGLSERLQATQKMRGK.....GKKGRNQNMKGMNLPF 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2267	99.7	455	AAU00910	S. aureus ffh prot
2	2261	99.4	455	22 AAU37473	Staphylococcus aur
3	2235	98.3	455	22 AAU33924	Staphylococcus aur
4	2235	98.3	455	22 AAU36535	Staphylococcus aur
5	1717.5	75.5	450	23 ABB49131	Listeria monocytog
6	1534	67.5	338	23 ABP39950	Staphylococcus epi
7	1418.5	62.4	472	22 AAU33449	Enterococcus faeca
8	1418.5	62.4	472	22 AAU35268	Enterococcus faeca
9	1315	57.8	518	23 ABB54962	Lactococcus lactis
10	1307	57.5	523	22 AAU37661	Streptococcus pneu

11	1306	57.4	466	22 AAU37932	Streptococcus pneu
12	1306	57.4	523	22 AAU01025	CFE 28 protein seq
13	1305	57.4	264	20 AAU00911	S. aureus ffh prot
14	1297	57.0	523	20 AAU99658	Streptococcus pneu
15	1297	57.0	523	23 AAU79163	Streptococcus pneu
16	1279.5	56.3	521	23 ABP26457	Streptococcus poly
17	1273.5	56.0	521	23 ABP26456	Streptococcus poly
18	1168	51.4	547	22 AAG92001	C glutamicum prote
19	1164.5	51.2	453	22 AAG34670	E. coli cellular p
20	1164.5	51.2	453	22 AAG98879	E. coli growth and
21	1150	50.6	457	22 AAU36355	Pseudomonas aerugi
22	1138	50.0	462	22 AAU35372	Haemophilus influe
23	1127.5	49.6	501	22 AAU38496	Salmonella typhi c
24	1014	44.6	531	22 AAU41288	Propionibacterium
25	869	38.2	448	20 AAU37201	Chlamydia trachoma
26	854	37.6	448	22 AAU35969	Helicobacter pylori
27	846	37.2	420	20 AAU81979	Ehrlichia sp. E74.
28	836	36.8	448	20 AAU34715	Chlamydia pneumoni
29	795	35.0	366	18 AAU20643	H. pylori cytoplasm
30	759.5	33.4	443	22 AAB95153	Putative P. abyssal
31	612.5	26.9	504	23 ABB57149	Mouse ischaemic co
32	592.5	26.1	535	17 AAU91310	Fungal signal reco
33	586	25.8	534	17 AAU91309	Fungal signal reco
34	576.5	25.4	508	22 ABB60834	Drosophila melanog
35	560	24.6	479	21 AAG41615	Arabidopsis thalia
36	548.5	24.1	463	21 AAG41616	Arabidopsis thalia
37	535.5	23.5	404	22 AAG74051	Human colon cancer
38	521	22.9	428	21 AAG41617	Arabidopsis thalia
39	432	19.0	497	22 AAU34785	E. coli cellular p
40	429.5	18.9	329	18 AAU11328	Bacillus subtilis
41	429	18.9	328	23 ABB48889	Listeria monocytog
42	428	18.8	204	18 AAU20559	Helicobacter pylori
43	428	18.8	204	18 AAU24688	H. pylori cytoplasm
44	428	18.8	332	22 AAB96695	Putative P. abyssal
45	426.5	18.8	415	23 ABP39972	Staphylococcus epi

#### ALIGNMENTS

RESULT 1  
AAU00910  
ID AAU00910 standard; Protein; 455 AA.  
XX AC AAU00910;  
XX DT 28-MAY-1999 (first entry)  
XX DE S. aureus ffh protein sequence.  
XX KW ffh gene; signal recognition particle; SRP; antimicrobial agent; vaccine;  
XX KW immunological response; gene therapy; infection; otitis media;  
XX KW conjunctivitis; toxic shock syndrome; septic arthritis.  
XX OS Staphylococcus aureus.  
XX PN EP902087-A2.  
XX PD 17-MAR-1999.  
XX PF 24-AUG-1998; 98EP-0306741.  
XX PR 10-SEP-1997; 97US-0927216.  
XX PR (SMIK ) SMITHKLINE BEECHAM.  
XX PI Wallis NG;  
XX DR WPI; 1999-169238/15.  
XX DR N-PSDB; AAX27221.  
XX PT New Staphylococcus aureus Signal Recognition Particle (SRP) with  
XX PT protein (ffh) and RNA (ffs) components - the SRP gene and protein

PT useful as diagnostic reagents and for prevention and treatment of  
PT Staphylococci infections which cause otitis media, septic arthritis  
PT and toxic shock syndrome  
XX Claim 11; Page 29-30; 35pp; English.  
XX This sequence is the Staphylococcus aureus signal recognition  
CC particle (SRP) fff component. Fff polynucleotides are useful for  
CC diagnosing a disease related to expression of fff polypeptides by  
CC analysing for the presence/amount of fff protein in a sample due to  
CC infection of a micro-organism with the gene, or determining the nucleic  
CC acid sequence encoding fff. Fff polypeptides and antagonists are useful  
CC for treatment of an individual in need (polypeptide) of, or needing to  
CC inhibit (antagonist) fff polypeptide levels. Fff polypeptides and  
CC polynucleotides are useful for identifying agonists and antagonists by  
CC binding and observing the affect of fff polypeptide activity, which are  
CC potential anti-microbial agents. Fff polypeptides and antigenic fragments  
CC are also useful for inducing an immunological response (T cell/antibody)  
CC to protect against disease, by direct administration (vaccine), or via a  
CC vector (gene therapy). Anti-fff antibodies are useful as antagonists, and  
CC for protecting against disease. Diseases diagnosed, prevented and treated  
CC include those caused by infection, especially bacterial infection,  
CC including otitis media, conjunctivitis, toxic shock syndrome, wound  
CC infection and septic arthritis. Fff polypeptides are antimicrobial, and  
CC are useful for bathing wounds and implants prior to surgical  
CC implantation.

XX Sequence 455 AA;

Query Match 99.7%; Score 2267; DB 20; Length 455;

Best Local Similarity 99.6%; Pred. No. 2.5e-174;  
Matches 454; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAFGLSERLOATMQMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60  
DB 1 MAFGLSERLOATMQMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60  
QY 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSTINMSNKPPTVVMVGLQAGKTTTACK 120  
DB 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSTINMSNKPPTVVMVGLQAGKTTTACK 120  
QY 121 LALLMRKKYNNKPMVAADIVRPAINOLQTVGKQIDIPVYSEGQVXPQOIVTNALKHA 180  
DB 121 LALLMRKKYNNKPMVAADIVRPAINOLQTVGKQIDIPVYSEGQVXPQOIVTNALKHA 180  
QY 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKKEIAPNEIMLVDSMTGQDAYNVAESFDDQ 240  
DB 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKKEIAPNEIMLVDSMTGQDAYNVAESFDDQ 240  
QY 241 LDVTGVTILKLDGTRGGAALSIRSVTOKPKFVGMSEKLDGLELPHPERMASRIILMGD 300  
DB 241 LDVTGVTILKLDGTRGGAALSIRSVTOKPKFVGMSEKLDGLELPHPERMASRIILMGD 300  
QY 301 VLSLTERAQQDVQEKAKDEKMPRESSFTLDDFLEQLDQVKNLGLDDIMKMPGNKM 360  
DB 301 VLSLTERAQQDVQEKAKDEKMPRESSFTLDDFLEQLDQVKNLGLDDIMKMPGNKM 360  
QY 361 KGLDLKNNSEKQIDHIKALIQSMTPAERNPDITLNSVRKKRIAGSGRSLQEVNRLMKQF 420  
DB 361 KGLDLKNNSEKQIDHIKALIQSMTPAERNPDITLNSVRKKRIAGSGRSLQEVNRLMKQF 420  
QY 421 NDMMKMMKQFTGGGKGGKGRNQMMNLKGNLPP 455  
DB 421 NDMMKMMKQFTGGGKGGKGRNQMMNLKGNLPP 455  
RESULT 2  
AAU37473  
ID AAU37473 standard; Protein; 455 AA.  
XX AAU37473;  
XX AAU37473;  
XX 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #1643.  
DE Antisense; prokaryotic cellular proliferation protein;  
XX antibiotic; antibacterial; drug design.  
KW Staphylococcus aureus.  
XX WO200170955-A2.  
XX 27-SEP-2001.  
XX 21-MAR-2001; 2001WO-US09180.  
XX 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI: 2001-611495/70.  
DR N-PSDB; AAS55332.  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX Example 3; Seq ID No 13066; 511pp; English.  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 455 AA;

Query Match 99.4%; Score 2261; DB 22; Length 455;

Best Local Similarity 99.3%; Pred. No. 7.6e-174;  
Matches 452; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAFGLSERLOATMQMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60  
DB 1 MAFGLSERLOATMQMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60  
QY 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSTINMSNKPPTVVMVGLQAGKTTTACK 120  
DB 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSTINMSNKPPTVVMVGLQAGKTTTACK 120  
QY 121 LALLMRKKYNNKPMVAADIVRPAINOLQTVGKQIDIPVYSEGQVXPQOIVTNALKHA 180  
DB 121 LALLMRKKYNNKPMVAADIVRPAINOLQTVGKQIDIPVYSEGQVXPQOIVTNALKHA 180

QY 181 KEEHDFVIIDTAGRLHIDEALMNEIKVEYKTAKEIPNEIMLVDSMTQDAVNVVAFSFDQ 240  
 DQ 181 KEEHDFVIIDTAGRLHIDEALMNEIKVEYKTAKEIPNEIMLVDSMTQDAVNVVAFSFDQ 240  
 QY 241 LDVTGVTLLKLDGTRGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASRILGMD 300  
 DQ 241 LDVTGVTLLKLDGTRGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASRILGMD 300  
 QY 301 VLSLIEKAQDVQDEKAKDLEKKMRSSFTLDDFLEQLDQVKNLGLPDDIMKMGMLPF 360  
 DQ 301 VLSLIEKAQDVQDEKAKDLEKKMRSSFTLDDFLEQLDQVKNLGLPDDIMKMGMLPF 360  
 QY 361 KGLDKLNMESEKQIDHIKAIQSMTPAERNPDTLVNSRKKRIAKGSGRSLOEVNRLMKQF 420  
 DQ 361 KGLDKLNMESEKQIDHIKAIQSMTPAERNPDTLVNSRKKRIAKGSGRSLOEVNRLMKQF 420  
 QY 421 NDMMKMMKQFTGGGKGKGRNOMNMLKGMNLPF 455  
 DQ 421 NDMMKMMKQFTGGGKGKGRNOMNMLKGMNLPF 455

RESULT 3  
 ID AAU33924 standard; Protein; 455 AA.  
 AC AAU33924;  
 XX  
 DT 14-FEB-2002 (first entry)  
 DE Staphylococcus aureus cellular proliferation protein #200.  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 OS Staphylococcus aureus.  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207272P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI: 2001-611495/70.  
 DR N-PSDB; AAS51783.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 5420; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: the sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 455 AA;

Query Match 98.3%; Score 2235; DB 22; Length 455;  
 Best Local Similarity 98.2%; Pred. No. 9.5e-172;  
 Matches 447; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAPEGLSERLOATQKMRGKGLTEADIKIMREVRLLALFEADVNEKVKYKFIKTVSRA 60  
 DQ 1 MAPEGLSERLOATQKMRGKGLTEADIKIMREVRLLALFEADVNEKVKYKFIKTVSRA 60  
 QY 61 LGSVMQSLTFGQOVIKIVODELTKLGGENTSNMGNKPTVVMVGLQGAGKTTTAKG 120  
 DQ 61 LGSVMQSLTFGQOVIKIVODELTKLGGENTSNMGNKPTVVMVGLQGAGKTTTAKG 120  
 QY 121 LALLMRKKYKPKMLVAADIYRPAALNQLQTVGKQIDIPVISEGQVKKPQIVTNALKHA 180  
 DQ 121 LALLMRKKYKPKMLVAADIYRPAALNQLQTVGKQIDIPVISEGQVKKPQIVTNALKHA 180  
 QY 181 KEEHDFVIIDTAGRLHIDEALMNEIKVEYKTAKEIPNEIMLVDSMTQDAVNVVAFSFDQ 240  
 DQ 181 KEEHDFVIIDTAGRLHIDEALMNEIKVEYKTAKEIPNEIMLVDSMTQDAVNVVAFSFDQ 240  
 QY 241 LDVTGVTLLKLDGTRGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASRILGMD 300  
 DQ 241 LDVTGVTLLKLDGTRGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASRILGMD 300  
 QY 301 VLSLIEKAQDVQDEKAKDLEKKMRSSFTLDDFLEQLDQVKNLGLPDDIMKMGMLPF 360  
 DQ 301 VLSLIEKAQDVQDEKAKDLEKKMRSSFTLDDFLEQLDQVKNLGLPDDIMKMGMLPF 360  
 QY 361 KGLDKLNMESEKQIDHIKAIQSMTPAERNPDTLVNSRKKRIAKGSGRSLOEVNRLMKQF 420  
 DQ 361 KGLDKLNMESEKQIDHIKAIQSMTPAERNPDTLVNSRKKRIAKGSGRSLOEVNRLMKQF 420  
 QY 421 NDMMKMMKQFTGGGKGKGRNOMNMLKGMNLPF 455  
 DQ 421 NDMMKMMKQFTGGGKGKGRNOMNMLKGMNLPF 455

RESULT 4  
 AAU36535  
 ID AAU36535 standard; Protein; 455 AA.  
 AC AAU36535;  
 XX  
 DT 14-FEB-2002 (first entry)  
 DE Staphylococcus aureus cellular proliferation protein #705.  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 OS Staphylococcus aureus.  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207272P.

```
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
DR WPI; 2001-611495/70.
DR N-PSDB; AAS4394.
XX
XX New polynucleotides for the identification and development of
PI antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12128; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 455 AA;
Query Match 98.3%; Score 2235; DB 22; Length 455;
Best Local Similarity 98.2%; Pred. No. 9.5e-172;
Matches 447; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAPEGLSERLOATMQMRGKGLTEADIKIMREVRLLAEADVNFVKVKEFIKTVSERA 60
Db 1 MAPEGLSERLOATMQMRGKGLTEADIKIMREVRLLAEADVNFVKVKEFIKTVSERA 60
QY 61 LGSQVMSLTPGQVQIKIVQDELTKLGGENTSIINNSKPPTVVMVGLQAGKTTAGK 120
Db 61 LGSQVMSLTPGQVQIKIVQDELTKLGGENTSIINNSKPPTVVMVGLQAGKTTAGK 120
QY 121 LALIMRKKYNKPMVAADYIPRAINQLOTVGKQIDIPVYSEGQVQKPOQIVTNALKHA 180
Db 121 LALIMRKKYNKPMVAADYIPRAINQLOTVGKQIDIPVYSEGQVQKPOQIVTNALKHA 180
QY 191 KEELHFDVIDTAGRLHIDEALMNEKEVEIAKPNELMVVDSMTGQDVAVNAESPDQ 240
Db 191 KEELHFDVIDTAGRLHIDEALMNEKEVEIAKPNELMVVDSMTGQDVAVNAESPDQ 240
QY 241 LDVTGVTLLKLDGTRFGGAALSISVTKPKIKFVGMSEKLDGLFHFPERNASRILMGD 300
Db 241 LDVTGVTLLKLDGTRFGGAALSISVTKPKIKFVGMSEKLDGLFHFPERNASRILMGD 300
QY 301 VLSLIEKAQODVDQEKADLEKKKRESFTLDDFLQOLDQVKNLGLPDDIMKPIGNMKM 360
Db 301 VLSLIEKAQODVDQEKADLEKKKRESFTLDDFLQOLDQVKNLGLPDDIMKPIGNMKM 360
QY 361 KGLDKLNMSKQIDHIIKAIQISMTPAERNPDFTLNSRKKRIAGSGRSQEVNRLMKQF 420
Db 361 KGLDKLNMSKQIDHIIKAIQISMTPAERNPDFTLNSRKKRIAGSGRSQEVNRLMKQF 420
QY 421 NDMMKKMMKQFTGGGKKGKGRNQNMMLKGNLPP 455
```

```
Db 421 NDMMKKMMKQFTGGGKKGKGRNQNMMLKGNLPP 455
```

```
RESULT 5
ABE49131
```

```
ID ABB49131 standard; Protein: 450 AA.
```

```
XX ABB49131;
```

```
AC ABB49131;
```

```
DT 05-FEB-2002 (first entry)
```

```
DE Listeria monocytogenes protein #1835.
```

```
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
```

```
XX Listeria monocytogenes.
```

```
XX WO200177335-A2.
```

```
XX 18-OCT-2001.
```

```
XX 11-APR-2001; 2001WO-FR01118.
```

```
XX 11-APR-2000; 2000FR-0004629.
```

```
XX (INSP ) INST PASTEUR.
```

```
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Bequerro F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehlend J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
XX WPI; 2002-010914/01.
```

```
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
```

```
PS Claim 6; SEQ ID No 1836; 192bp; French.
```

```
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC polymorphisms and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies. Identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
```

```
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX Sequence 450 AA;
```

```
Query Match 75.5%; Score 1717.5; DB 23; Length 450;
```

```
Best Local Similarity 72.7%; Pred. No. 4.8e-130;
```

```
Matches 331; Conservative 64; Mismatches 55; Indels 5; Gaps 1;
```

```
QY 1 MAPEGLSERLOATMQMRGKGLTEADIKIMREVRLLAEADVNFVKVKEFIKTVSERA 60
|||||: ||| |||:|||||: |||:| ||||||| |||||||:|||||
```



Db 1 MAFEGLAGRQETMNTKRGKVNADYKEMMREVRLLALEADVNFVKVQKFIKTVSERA 60  
QY 61 LGSVMQSLTPGOQVIKIVQDELTKLMGGENTSIINMSNKPPTVVMVGLQAGKTTTAGK 120  
Db 61 VQADYKMSLTPGOQVIKIVQDELTKLMGGESKIGTADRPPTVVMVGLQAGKTTTSGK 120  
QY 121 LALLMRKYNKPMVAADYIRPAAINOLQTVGKQIDIPYSEGQVFPQOIVTNALKHA 180  
Db 121 LALLMRKYNKPMVAADYIRPAAINOLQTVGKQIDIPYSEGQVFPQOIVTNALKHA 180  
QY 181 KEEHLDVFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVDSMTGQDANVAESFDDQ 240  
Db 181 KEEHLDVFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVDSMTGQDANVAESFDDQ 240  
QY 241 LDVTGVTITKLDGTRGGAALSIRSVTQKPIKFGMSEKLDGLELPHPERMASRILGMD 300  
Db 241 LEITGVVITKLDGTRGGAALSIRSVTQKPIKFGMSEKLDGLELPHPERMASRILGMD 300  
QY 301 VLSLIEKAQDDVDQEKAKDLEKKMRESSFTLDDFLEOLDQVKNLGPLDDIMKMPGNKM 360  
Db 301 VLSLIEKAQDDVDQEKAKDLEKKMRESSFTLDDFLEOLDQVKNLGPLDDIMKMPGNKM 360  
QY 361 KGLDKLNMSKOIDHIAIKISQMTPAERNPDTLVNRSKRIAKSGRSLOEVNRLAKQF 420  
Db 361 KGLDNVYDDQOLGHEIAIKISQMTKNEKNDPDIINASRRKRIARGSGRPVQEIINELLKQF 420  
QY 421 NDMMKMKOFFGGGKGKGRNOMQNLKGMNLPF 455  
Db 421 AEMKKMKOMTGGGKGKGRK-----NPPGNFKMPF 450

RESULT 6  
ABP39950  
ID ABP39950 standard; Protein; 338 AA.  
XX AC ABP39950;  
XX DT 24-JUL-2002 (first entry)  
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4795.  
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX KW antibacterial; gene therapy.  
XX OS Staphylococcus epidermidis.  
XX PN US6380370-B1.  
XX PD 30-APR-2002.  
XX PF 13-AUG-1998; 98US-0134001.  
XX PR 14-AUG-1997; 97US-055779P.  
XX PR 08-NOV-1997; 97US-064964P.  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX PI Doucette-Stamm LA, Bush D;  
XX XX WPI; 2002-381255/41.  
XX DR N-PSDB; ABN92495.  
XX XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
PT Disclosure; SEQ ID 4795; 267pp; English.  
XX XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37950. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to

CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX SQ Sequence 338 AA;  
Query Match 67.5%; Score 1534; DB 23; Length 338;  
Best Local Similarity 92.5%; Pred. No. 2e-115;  
Matches 307; Conservative 16; Mismatches 9; Indels 0; Gaps 0;  
QY 1 MAFEGLAGRQETMNTKRGKVNADYKEMMREVRLLALEADVNFVKVQKFIKTVSERA 60  
Db 6 MAFEGLAGRQETMNTKRGKVNADYKEMMREVRLLALEADVNFVKVQKFIKTVSERA 65  
QY 61 LGSVMQSLTPGOQVIKIVQDELTKLMGGENTSIINMSNKPPTVVMVGLQAGKTTTAGK 120  
Db 66 LGSVMQSLTPGOQVIKIVQDELTKLMGGENTSIINMSNKPPTVVMVGLQAGKTTTAGK 125  
QY 121 LALLMRKYNKPMVAADYIRPAAINOLQTVGKQIDIPYSEGQVFPQOIVTNALKHA 180  
Db 126 LALLMRKYNKPMVAADYIRPAAINOLQTVGKQIDIPYSEGQVFPQOIVTNALKHA 185  
QY 181 KEEHLDVFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVDSMTGQDANVAESFDDQ 240  
Db 186 KEEHLDVFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVDSMTGQDANVAESFDDQ 245  
QY 241 LDVTGVTITKLDGTRGGAALSIRSVTQKPIKFGMSEKLDGLELPHPERMASRILGMD 300  
Db 246 LDVTGVTITKLDGTRGGAALSIRSVTQKPIKFGMSEKLDGLELPHPERMASRILGMD 305  
QY 301 VLSLIEKAQDDVDQEKAKDLEKKMRESSFTLDD 332  
Db 306 VLSLIEKAQDDVDQEKAKDLEKKMRESSFTLDD 337  
RESULT 7  
AAU33449  
ID AAU33449 standard; Protein; 472 AA.  
XX AC AAU33449;  
XX DT 14-FEB-2002 (first entry)  
XX DE Enterococcus faecalis cellular proliferation protein #85.  
XX KW Antisense; prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.  
XX OS Enterococcus faecalis.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US09180.  
XX PR 21-MAR-2000; 2000US-191078P.  
XX PR 23-MAY-2000; 2000US-206848P.  
XX PR 26-MAY-2000; 2000US-207272P.  
XX PR 23-OCT-2000; 2000US-242578P.  
XX PR 27-NOV-2000; 2000US-253625P.  
XX PR 22-DEC-2000; 2000US-257931P.  
XX PR 16-FEB-2001; 2001US-269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlens KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
XX DR N-PSDB; AAS51308.

XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX  
PS Example 3; Seq ID No 4945; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 472 AA;

Query Match 62.4%; Score 1418.5; DB 22; Length 472;  
Best Local Similarity 59.0%; Pred. No. 6.5e-106;  
Matches 273; Conservative 86; Mismatches 89; Indels 15; Gaps 2;

QY 1 MAFEGLSERLOATMOKMRGKGLTADIKIMREVELALFEADVNFVKVKEFIKTVSERA 60  
DB 1 MAFESLNRLOQAMSKIRKRGKGVADYKEMREIRLALLEADVNLQVVKDFTRVRERA 60  
QY 61 LGSVMQSLPFGQOVIKIQDELTKLGGENTSSINMSNKPPTVVMVGLQAGKTTAGK 120  
DB 61 VGVVLESLSPAQOVIKIVDEELTKLGSFVELNKSPIPTVIMTGLQAGKTTFTCK 120  
QY 121 LALLMRKYNKPKMVAADIVRPAINQLQTVGKQIDIPYSEGQVXPQQIVYNALKHA 180  
DB 121 LAKHLMKTENARPLIAGDVYRPAADQLKVLGQLEVPFDMGTDPANPVEIVRQGLALA 180  
QY 181 KEEHLDFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVVDSMTGDAVNVAESFDDQ 240  
DB 181 KEKKNDYVIDTAGRLHIDEALMDELKOIKELANPEILLVVDATGQDAVNVAESFNEQ 240  
QY 241 LDVTGVTKLDGTRGGAALSRVTKQPKFVGMSEKLDGLELPHPERVAKSILGMD 300  
DB 241 LGITGVWTKMDGTRGGAALSIRAVTGAPIKFVGSGETDLEIFHPDRMSSRILGMD 300  
QY 301 VLSLEKAOQVDQERAKLEKMRSEFTLDDFLEQLDDVKNLGPLDDIMKMPGNKM 360  
DB 301 MTLIEKAQDYDEKKAELAAQKNSENFDFNDFLEQLDDVGMGPFIEDLLKMTIPGNSM 360  
QY 361 KGLDKLNMSEKQIDHITKAIQSTPAERNPDTLNVSRRKRIAGSGRSLOEVNRMKQF 420  
DB 361 PGIENVKVPDPRDVARKAWVLSMTPAERENPDLLNPSRRRRRIAAGSGNSVVEVNRMIKQF 420  
QY 421 NDMMKMKQFT-----GGGKGGKGNQNMK 449  
DB 421 KESKMMQOMSGMDNIPGMDQMLGGGVKGLGKM-AMNRMK 462

RESULT 8  
AAU35268  
ID AAU35268 standard; Protein; 477 AA.  
XX  
AC AAU35268;  
XX  
QT 14-FEB-2002 (first entry)  
XX

DE Enterococcus faecalis cellular proliferation protein #555.  
XX Antisense; prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.  
XX  
XX Enterococcus faecalis.  
XX WO200170955-A2.  
PD 27-SEP-2001.  
PF 21-MAR-2001; 2001WO-US09180.  
XX 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX (ELIT-) ELITRA PHARM INC.  
PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
XX WPI; 2001-611495/70.  
DR N-PSDB; AAS53127.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Example 3; Seq ID No 10861; 511pp; English.  
PS The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 477 AA;

Query Match 62.4%; Score 1418.5; DB 22; Length 477;  
Best Local Similarity 59.0%; Pred. No. 6.6e-106;  
Matches 273; Conservative 86; Mismatches 89; Indels 15; Gaps 2;

QY 1 MAFEGLSERLOATMOKMRGKGLTADIKIMREVELALFEADVNFVKVKEFIKTVSERA 60  
DB 1 MAFESLNRLOQAMSKIRKRGKGVADYKEMREIRLALLEADVNLQVVKDFTRVRERA 60  
QY 61 LGSVMQSLPFGQOVIKIQDELTKLGGENTSSINMSNKPPTVVMVGLQAGKTTAGK 120  
DB 61 VGVVLESLSPAQOVIKIVDEELTKLGSFVELNKSPIPTVIMTGLQAGKTTFTCK 120  
QY 121 LALLMRKYNKPKMVAADIVRPAINQLQTVGKQIDIPYSEGQVXPQQIVYNALKHA 180  
DB 121 LAKHLMKTENARPLIAGDVYRPAADQLKVLGQLEVPFDMGTDPANPVEIVRQGLALA 180  
QY 181 KEEHLDFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVVDSMTGDAVNVAESFDDQ 240



CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 523 AA;

Query Match 57.5%; Score 1307; DB 22; Length 523;

Best Local Similarity 55.1%; Pred. No. 7.3e-37;

Matches 253; Conservative 90; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGLSRLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVEFIKTVSERA 60  
 DB 1 MAFESLTERLQNFKNLRKKGKISDSVQEQATKEIRLALLEADVALPVVDKIKYRERA 60  
 QY 61 LGSVYMSLTPGQVQIKIVQDELTKLMGENTSNMKNKPTTVMMVGLQGAGKTTTAGK 120  
 DB 61 VGHEVIDTLNPAQOIIKIVDEELTAVLGSDAEIKSPKIPTIMMVGLOGAGKTTTAGK 120  
 QY 121 LALLMRKKYKPKMLVAADIRPAALNQLQVQKIDIPVYSEGDOVKPQQIVTNALKHA 180  
 DB 121 LANKLKKEENARPLMVAADIRPAALDQLKILGQIDVPVFGTVEPAVEIVRQGLEQA 180  
 QY 181 KEEHLDFVIIDTAGRLHIDEALMNLKELKEKTAKEPNEIMLVDSMTGQDANVVAESDDQ 240  
 DB 181 QTNHNDYVLIDTAGRLQIDELLNLRDVKALQAQNEILLVDMITGQEAANVAREFNAQ 240  
 QY 241 LDVTGVTLTKLDGTRGAALSTRVTPKPIKFGVMSKELDLGLELPHPERMASRLTGMGD 300  
 DB 241 LEVTGVTLTKLDGTRGAALSVRHTGPKIKFTGTEKITDIETHPDRMSSRLTGMGD 300  
 QY 301 VLSLIEKAQDQDQEKAKLEKKMRSSFTLDDFLEQLDQVKNLGPDLDDIMKMPGNMKM 360  
 DB 301 MLTLIEKASQEDQKALENAEKMRNTDFDNFDQLDQVQNGPMDLLAKMIFGMANN 360  
 QY 361 KGLDLKNMSEKQIDHIKAIQSWTPAERNNDPTLVNSKKRIAKGGRSLQEVNRLMKQF 420  
 DB 361 PALQNMKYDERQIARKRAIVSSMTPBERENPDLLNPSRRRIAGSGNTFVEVNFKIDF 420  
 QY 421 NDMKKMKQFTGGCGKKGKKNQMNKMG-----NLP 454  
 DB 421 NQAKOLMGVMSG-----DANKMKMGKMGINDNNLP 450

RESULT 11

ID AAU37932

XX AAU37932 standard; Protein; 466 AA.

AC AAU37932;

XX AAU37932;

DT 14-FEB-2002 (first entry)

XX Streptococcus pneumoniae cellular proliferation protein #361.

XX Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX Streptococcus pneumoniae.

OS

XX WO200170955-A2.  
 PN 27-SEP-2001.  
 XX 21-MAR-2001; 2001WO-US09180.  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242378P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-369308P.  
 XX (ELIT-) ELITPA PHARM INC.  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-6111495/70.  
 DR N-PSDB; AAS55791.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX Example 3; Seq ID No 13525; 511pp; English.  
 PS The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 466 AA;

Query Match 57.4%; Score 1306; DB 22; Length 466;

Best Local Similarity 54.9%; Pred. No. 7.5e-97;

Matches 252; Conservative 91; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGLSRLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVEFIKTVSERA 60  
 DB 1 MAFESLTERLQNFKNLRKKGKISDSVQEQATKEIRLALLEADVALPVVDKIKYRERA 60  
 QY 61 LGSVYMSLTPGQVQIKIVQDELTKLMGENTSNMKNKPTTVMMVGLQGAGKTTTAGK 120  
 DB 61 VGHEVIDTLNPAQOIIKIVDEELTAVLGSDAEIKSPKIPTIMMVGLOGAGKTTTAGK 120  
 QY 121 LALLMRKKYKPKMLVAADIRPAALNQLQVQKIDIPVYSEGDOVKPQQIVTNALKHA 180  
 DB 121 LANKLKKEENARPLMVAADIRPAALDQLKILGQIDVPVFGTVEPAVEIVRQGLEQA 180  
 QY 181 KEEHLDFVIIDTAGRLHIDEALMNLKELKEKTAKEPNEIMLVDSMTGQDANVVAESDDQ 240  
 DB 181 QTNHNDYVLIDTAGRLQIDELLNLRDVKALQAQNEILLVDMITGQEAANVAREFNAQ 240  
 QY 241 LDVTGVTLTKLDGTRGAALSTRVTPKPIKFGVMSKELDLGLELPHPERMASRLTGMGD 300  
 DB 241 LEVTGVTLTKLDGTRGAALSVRHTGPKIKFTGTEKITDIETHPDRMSSRLTGMGD 300



PT Staphylococci infections which cause otitis media, septic arthritis  
 PT and toxic shock syndrome  
 XX  
 PS Claim 11; Page 31-32; 35pp; English.  
 XX  
 CC This sequence is the Staphylococcus aureus signal recognition  
 CC particle (SRP) ffh component. Ffh polynucleotides are useful for  
 CC diagnosing a disease related to expression of ffh polypeptides by  
 CC analysing for the presence/amount of ffh protein in a sample due to  
 CC infection of a micro-organism with the gene, or determining the nucleic  
 CC acid sequence encoding ffh. Ffh polypeptides and antagonists are useful  
 CC for treatment of an individual in need (polypeptide) of, or needing to  
 CC inhibit (antagonist) ffh polypeptide levels. Ffh polypeptides and  
 CC polynucleotides are useful for identifying agonists and antagonists by  
 CC binding and observing the affect of ffh polypeptide activity, which are  
 CC potential anti-microbial agents. Ffh polypeptides and antigenic fragments  
 CC are also useful for inducing an immunological response (T cell/antibody)  
 CC to protect against disease, by direct administration (vaccine), or via a  
 CC vector (gene therapy). Anti-fhh antibodies are useful as antagonists, and  
 CC for protecting against disease. Diseases diagnosed, prevented and treated  
 CC include those caused by infection, especially bacterial infection,  
 CC including otitis media, conjunctivitis, toxic shock syndrome, wound  
 CC infection and septic arthritis. Ffh polypeptides are antimicrobial, and  
 CC are useful for bathing wounds and implants prior to surgical  
 CC implantation.  
 XX  
 SQ Sequence 264 AA;  
 Query Match 57.4%; Score 1305; DB 20; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-97; Mismatches 0; Gaps 0;  
 Matches 264; Conservative 0; Indels 0;  
 QY 1 MAFEGLSERLQATMQKRGKGLTEADIKIMREVRALFEADYNFKVKEFIKTVSERA 60  
 DB 1 MAFEGLSERLQATMQKRGKGLTEADIKIMREVRALFEADYNFKVKEFIKTVSERA 60  
 QY 61 LGSVDVMSLTPGQGVKIVQDELTKLMGGENTSINNSKNKPTVVMVGLQGAGKTTAGK 120  
 DB 61 LGSVDVMSLTPGQGVKIVQDELTKLMGGENTSINNSKNKPTVVMVGLQGAGKTTAGK 120  
 QY 121 LALLMRKKNKPMPLVAADIYRPAAINQLOTGVKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
 DB 121 LALLMRKKNKPMPLVAADIYRPAAINQLOTGVKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
 QY 181 KEEHDFVIIDTAGRLHIDEALMNEELKEVKEIAKPNEIMLVYDSMTGQDAVNVASFDDQ 240  
 DB 181 KEEHDFVIIDTAGRLHIDEALMNEELKEVKEIAKPNEIMLVYDSMTGQDAVNVASFDDQ 240  
 QY 241 LDVTVGTLTKLGDTRGGGALSIR 264  
 DB 241 LDVTVGTLTKLGDTRGGGALSIR 264  
 RESULT 14  
 AAW99658  
 ID AAW99658 standard; Protein; 523 AA.  
 XX  
 AC AAW99658;  
 XX  
 DT 21-MAY-1999 (first entry)  
 XX  
 DE Streptococcus pneumoniae ffh protein.  
 XX  
 KW Streptococcus pneumoniae; ffh; fifty-four homologue; antibacterial;  
 KW infection; otitis media; conjunctivitis; bacteraemia; sinusitis;  
 KW pleural empyema; endocarditis; meningitis.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN EP900843-A2.  
 XX  
 XX 10-MAR-1999.  
 PD  
 XX

PF 20-AUG-1998; 98EP-0306685.  
 XX  
 PR 02-SEP-1997; 97US-0923772.  
 XX  
 PA (SMIX ) SMITHKLINE BEECHAM.  
 XX  
 PI Black WT;  
 XX  
 DR WPI; 1999-155936/14.  
 DR N-PSDB; AAX19484.  
 XX  
 PT New Streptococcus pneumoniae Fifty-Four Homologue (Ffh) polypeptide  
 PT and polynucleotide - useful as diagnostic reagents and for  
 PT prevention and treatment of Streptococci infections, which cause  
 PT otitis media, sinusitis and conjunctivitis  
 XX  
 CC Claim 12; Page 6; 21pp; English.  
 XX  
 CC The present sequence represents the Streptococcus pneumoniae fifty-four  
 CC homologue (ffh) protein, which is a component of the protein secretory  
 CC apparatus in bacteria, and the bacterial homologue of the eukaryotic  
 CC Signal Recognition Particle. Ffh proteins and polynucleotides are useful  
 CC for diagnosing diseases related to over or underexpression of ffh protein  
 CC by identifying mutations in the ffh gene, or determining ffh protein or  
 CC mRNA expression levels due to an infection of an organism with the ffh  
 CC gene. They can diagnose the stage and type of infection. Ffh proteins are  
 CC also useful for screening for compounds which affect activity of the  
 CC stimulation or inhibition of the binding to ffh protein and observing the  
 CC treatment to inhibit (antagonist) i.e. antibacterial drugs) or enhance  
 CC (agonist) ffh activity, in addition to direct administration of ffh  
 CC proteins to treat conditions associated with a lack of ffh protein, or  
 CC direct administration of antisense sequences to prevent expression. Ffh  
 CC proteins (administered directly, in a vector and as a vaccine) and  
 CC antibodies induce an immune response to immunise and prevent disease.  
 CC Diseases diagnosed, prevented or treated include: bacterial infections,  
 CC especially Streptococcus pneumoniae infections, which cause otitis media,  
 CC conjunctivitis, bacteraemia, sinusitis, pleural empyema, endocarditis and  
 CC especially meningitis. Ffh proteins, polynucleotides and their  
 CC (ant)agonists can prevent adhesion of bacteria to matrix proteins, and  
 CC are useful for use on wounds and body implants to prevent bacterial  
 CC infection.  
 XX  
 SQ Sequence 523 AA;  
 Query Match 57.0%; Score 1297; DB 20; Length 523;  
 Best Local Similarity 54.7%; Pred. No. 4.7e-96;  
 Matches 251; Conservative 91; Mismatches 103; Indels 14; Gaps 2;  
 QY 1 MAFEGLSERLQATMQKRGKGLTEADIKIMREVRALFEADYNFKVKEFIKTVSERA 60  
 DB 1 MAFEGLSERLQATMQKRGKGLTEADIKIMREVRALFEADYNFKVKEFIKTVSERA 60  
 QY 61 LGSVDVMSLTPGQGVKIVQDELTKLMGGENTSINNSKNKPTVVMVGLQGAGKTTAGK 120  
 DB 61 LGSVDVMSLTPGQGVKIVQDELTKLMGGENTSINNSKNKPTVVMVGLQGAGKTTAGK 120  
 QY 121 LALLMRKKNKPMPLVAADIYRPAAINQLOTGVKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
 DB 121 LALLMRKKNKPMPLVAADIYRPAAINQLOTGVKQIDIPVYSEGQVKKPQOIVTNALKHA 180  
 QY 181 KEEHDFVIIDTAGRLHIDEALMNEELKEVKEIAKPNEIMLVYDSMTGQDAVNVASFDDQ 240  
 DB 181 KEEHDFVIIDTAGRLHIDEALMNEELKEVKEIAKPNEIMLVYDSMTGQDAVNVASFDDQ 240  
 QY 241 LDVTVGTLTKLGDTRGGGALSIR 264  
 DB 241 LDVTVGTLTKLGDTRGGGALSIR 264  
 QY 301 VLSLIEKAQDVQDEKAKLEKXMRSSFTLDDFLEQLDQVKNLGLPDDIMKMPGMNM 360  
 DB 301 VLSLIEKAQDVQDEKAKLEKXMRSSFTLDDFLEQLDQVKNLGLPDDIMKMPGMNM 360

1 MAFESLTERIONVFKNLRKKGKISESDVOEATKEIRLALLEADVALPVKDFIKKVRERA 60

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:17:12 ; Search time 18 seconds  
(without alignments)  
743.746 Million cell updates/sec

Title: US-09-943-108A-2  
Perfect score: 2274  
Sequence: 1 MAFEGLSERLOATMOKMRGK.....GKKGRNQNMKGMNLPF 455

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2274	100.0	455	4	US-09-035-382-2
2	1534	67.5	338	4	US-09-134-001C-4795
3	1305	57.4	264	4	US-09-035-382-4
4	1297	57.0	523	2	US-08-923-772-2
5	1297	57.0	523	4	US-09-385-287-2
6	846	37.2	420	4	US-09-066-047-8
7	586	25.8	534	2	US-08-317-401E-2
8	578	25.4	552	2	US-08-317-401E-4
9	429.5	18.9	329	4	US-08-981-527A-8
10	426.5	18.8	415	4	US-09-134-001C-4817
11	421	18.5	416	4	US-09-007-476-2
12	397.5	17.5	425	2	US-08-986-963-2
13	397.5	17.5	430	3	US-09-007-484-2
14	397.5	17.5	430	4	US-09-309-682-2
15	390	17.2	277	3	US-09-007-484-4
16	390	17.2	277	4	US-09-309-682-4
17	204.5	9.0	105	4	US-08-981-527A-13
18	179.5	7.9	105	4	US-08-981-527A-12
19	149.5	6.6	129	4	US-08-981-527A-11
20	136	6.0	606	4	US-08-477-831C-2
21	135.5	6.0	631	4	US-08-477-831C-11
22	132	5.8	1010	4	US-09-134-001C-5178
23	131	5.8	10182	4	US-09-134-001C-3159
24	130.5	5.7	1786	4	US-08-973-462-8
25	130	5.7	3696	4	US-09-134-001C-5080
26	127.5	5.6	1093	4	US-09-315-793-52
27	124	5.5	878	4	US-09-134-001C-4378

28	123	5.4	118	4	US-08-981-527A-10	Sequence 10, Appli
29	122.5	5.4	800	3	US-08-776-265-3	Sequence 3, Appli
30	119.5	5.3	423	2	US-08-250-731C-10	Sequence 10, Appli
31	119.5	5.3	1319	2	US-08-290-731C-2	Sequence 2, Appli
32	119.5	5.3	1336	2	US-08-290-731C-6	Sequence 6, Appli
33	119	5.2	593	2	US-08-591-079-8	Sequence 8, Appli
34	118.5	5.2	512	4	US-09-134-001C-4349	Sequence 4349, Ap
35	118.5	5.2	728	4	US-09-134-001C-4968	Sequence 4968, Ap
36	118	5.2	2482	1	US-08-328-254-6	Sequence 6, Appli
37	118	5.2	3248	1	US-08-353-700-1	Sequence 1, Appli
38	118	5.2	3248	5	PCT-US95-16216-1	Sequence 1, Appli
39	117	5.1	480	2	US-08-913-477-17	Sequence 17, Appli
40	116.5	5.1	573	4	US-08-235-836C-112	Sequence 112, App
41	116	5.1	2285	4	US-09-308-375-2	Sequence 2, Appli
42	115.5	5.1	454	4	US-09-134-001C-3547	Sequence 3547, Ap
43	115.5	5.1	804	4	US-09-134-001C-5218	Sequence 5218, Ap
44	115.5	5.1	1886	4	US-08-938-105-3	Sequence 3, Appli
45	115.5	5.1	1939	4	US-09-310-187A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-035-382-2  
; Sequence 2, Application US/09035382  
; Patent No. 6284515  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael T.  
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES  
; TITLE OF INVENTION: AND POLYNUCLEOTIDES  
; FILE REFERENCE: GMS0035  
; CURRENT APPLICATION NUMBER: US/09/035,382  
; CURRENT FILING DATE: 1998-03-05  
; EARLIER APPLICATION NUMBER: 60/057,890  
; EARLIER FILING DATE: 1997-09-03  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 455  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-035-382-2

Query Match Similarity 100.0%; Score 2274; DB 4; Length 455;  
Best Local Similarity 100.0%; Pred. No. 1.2e-190;  
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAFEGLSERLOATMOKMRGKGLTEADIKIMMRVRLALFEADVNFVKVKEFIKTVSERA	60
Db	1	MAFEGLSERLOATMOKMRGKGLTEADIKIMMRVRLALFEADVNFVKVKEFIKTVSERA	60
Qy	61	LGSDVMQSLTPGQGVKIVQDELTKLMGENTSINMSNKPTVVMVGLQAGKTTAGK	120
Db	61	LGSDVMQSLTPGQGVKIVQDELTKLMGENTSINMSNKPTVVMVGLQAGKTTAGK	120
Qy	121	LALLMKRYNKKPMLVAADIRPAAINQLOVWGKIDIPVYSEGQVKPQOIVTNALKHA	180
Db	121	LALLMKRYNKKPMLVAADIRPAAINQLOVWGKIDIPVYSEGQVKPQOIVTNALKHA	180
Qy	181	KEEHLDFVIIDTAGRLHIDPALMNEKVEKIAKPNELMVLVDSMTGDAVNAESDDQ	240
Db	181	KEEHLDFVIIDTAGRLHIDPALMNEKVEKIAKPNELMVLVDSMTGDAVNAESDDQ	240
Qy	241	LDVGTGVTITKLDGTRGGAALSIRSVTKPKIFVGMSEKLDGLFHPERWASRLGMGD	300
Db	241	LDVGTGVTITKLDGTRGGAALSIRSVTKPKIFVGMSEKLDGLFHPERWASRLGMGD	300
Qy	301	VLSLIEKAQDDQDEKAKLEKKMRESSFTLDDFLEQLDQVKNLGPLDDIMKMPGNMKM	360
Db	301	VLSLIEKAQDDQDEKAKLEKKMRESSFTLDDFLEQLDQVKNLGPLDDIMKMPGNMKM	360
Qy	361	KGLDKLNMSKQIDHIKAIIOQSMTFAERNPDITLVNSKRIAKGSGRSLQEVRLMKQF	420



```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-035-382-4

Query Match          57.4%; Score 1305; DB 4; Length 264;
Best local similarity 100.0%; Pred. No. 2.5e-106;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 WAFGLSERLQATMGKRGKGLTEADIKIMRVRLLALFEADVNFVKVEFIKTVSERA 60
   |||
DB 1 WAFGLSERLQATMGKRGKGLTEADIKIMRVRLLALFEADVNFVKVEFIKTVSERA 60

QY 61 LGSVMQSLTPGQQVQIKIVODELTKLMGGENTSINMSNKPPTVVMVGLQAGKTTTAK 120
   |||
DB 61 LGSVMQSLTPGQQVQIKIVODELTKLMGGENTSINMSNKPPTVVMVGLQAGKTTTAK 120

QY 121 LALLMRKKYKMPMLVADIRPAAINQLQTVGKQIDIPVSEGDQVKPQOIYVNALKHA 180
   |||
DB 121 LALLMRKKYKMPMLVADIRPAAINQLQTVGKQIDIPVSEGDQVKPQOIYVNALKHA 180

QY 181 KEEHLDEVIIDTAGRLHIDEALMNLKEVKEIAKPNELVLVDSMTGDAVNVAESFDDQ 240
   |||
DB 181 KEEHLDEVIIDTAGRLHIDEALMNLKEVKEIAKPNELVLVDSMTGDAVNVAESFDDQ 240

QY 241 LDVTGVTLLKLDGDRGGAALSIR 264
   |||
DB 241 LDVTGVTLLKLDGDRGGAALSIR 264

RESULT 4
US-08-923-772-2
; Sequence 2, Application US/08923772
; Patent No. 5972651
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: NOVEL ffh
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,772
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

US-08-923-772-2

```
Query Match          57.0%; Score 1297; DB 2; Length 523;
Best Local Similarity 54.7%; Pred. No. 3.4e-105;
Matches 251; Conservative 91; Mismatches 103; Indels 14; Gaps 2;

QY 1 MAFEGLSERLQATMGKMGKGLTEADIKIMREVRLLALFEADYVNFVKVKEFIKTVSERA 60
DB 1 MAFESUTERLQNVFNKLRKKGISDQVQATKEIRLALLEADVALPVVKDFIKKVERA 60

QY 61 LGSVDWQSLTPGQOVIKIVQDELTKLMGENTSIINMSKPTVVMVGLQAGKTTAGK 120
DB 61 VGEHVIDTLNPAQOIKIVDELTAVLGSDTAEIISKPKITIIIMVGLQAGKTTAGK 120

QY 121 LALLMRKYNKPMVAADIIYRPAAINQLQTVGKQIDIPVYSEGQVQKQOIVTNALKHA 180
DB 121 LANKLKEENARPLMTAADIIYRPAIDQLKTLGQOIVPVFALGTVEVPAVEIVRQGLEQA 180

QY 181 KEHLDFVIDTAGRLHIDEALMNEKKEVKEIAKPNELIMVDSMTGQDANVAESFDDQ 240
DB 181 QTNHNDYVLDITAGRLQIDELLMNELRDVKVLAQNEILLVVDAMIGQEAANVAREFNAQ 240

QY 241 LDVTGVILTKIDGTRGGAALSIKRSVTKQPIKFVGMSEKLDGLSLFHPERMASSRIILGMD 300
DB 241 LEVTGVILTKIDGTRGGAALSVRHITGPKIKFTGTGKIDTIDTFPHDRMSSRIILGMD 300

QY 301 VLSLIEKAQDVQDEKADLEKKMRESSFTLDDFLEQLDQVKNLGLPDDIMKIPGMNKM 360
DB 301 MLTLIEKASQYDEQKALEMAEKRENTDFDNFDIDQDVQVQNMGMEDLLKMIPEGANN 360

QY 361 KGLDKLNMSEKQIDHIAKIQSTWPAERNPDTLVNSRKKRIAKGSRSLQEVNRLMKQF 420
DB 361 PALQNMKYDERQIARKRAIVSSMTSERENPDLLNPSRRRIAGSGNTFVEVNFKIDF 420

QY 421 NDMMKMKKQFTGGGKGGKGRNQNMQLKGM-----NLP 454
DB 421 NOAKQLMQGVMSG-----DMNKKMKQMGINPNLP 450
```

## RESULT 5

```
US-09-385-287-2
; Sequence 2, Application US/09385287
; Patent No. 6350857
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: NOVEL fff
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,287
; FILING DATE: 30-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/923,772
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
```

TELEX: &lt;Unknown&gt;

```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-385-287-2
```

```
Query Match          57.0%; Score 1297; DB 4; Length 523;
Best Local Similarity 54.7%; Pred. No. 3.4e-105;
Matches 251; Conservative 91; Mismatches 103; Indels 14; Gaps 2;

QY 1 MAFEGLSERLQATMGKMGKGLTEADIKIMREVRLLALFEADYVNFVKVKEFIKTVSERA 60
DB 1 MAFESUTERLQNVFNKLRKKGISDQVQATKEIRLALLEADVALPVVKDFIKKVERA 60

QY 61 LGSVDWQSLTPGQOVIKIVQDELTKLMGENTSIINMSKPTVVMVGLQAGKTTAGK 120
DB 61 VGEHVIDTLNPAQOIKIVDELTAVLGSDTAEIISKPKITIIIMVGLQAGKTTAGK 120

QY 121 LALLMRKYNKPMVAADIIYRPAAINQLQTVGKQIDIPVYSEGQVQKQOIVTNALKHA 180
DB 121 LANKLKEENARPLMTAADIIYRPAIDQLKTLGQOIVPVFALGTVEVPAVEIVRQGLEQA 180

QY 181 KEHLDFVIDTAGRLHIDEALMNEKKEVKEIAKPNELIMVDSMTGQDANVAESFDDQ 240
DB 181 QTNHNDYVLDITAGRLQIDELLMNELRDVKVLAQNEILLVVDAMIGQEAANVAREFNAQ 240

QY 241 LDVTGVILTKIDGTRGGAALSIKRSVTKQPIKFVGMSEKLDGLSLFHPERMASSRIILGMD 300
DB 241 LEVTGVILTKIDGTRGGAALSVRHITGPKIKFTGTGKIDTIDTFPHDRMSSRIILGMD 300

QY 301 VLSLIEKAQDVQDEKADLEKKMRESSFTLDDFLEQLDQVKNLGLPDDIMKIPGMNKM 360
DB 301 MLTLIEKASQYDEQKALEMAEKRENTDFDNFDIDQDVQVQNMGMEDLLKMIPEGANN 360

QY 361 KGLDKLNMSEKQIDHIAKIQSTWPAERNPDTLVNSRKKRIAKGSRSLQEVNRLMKQF 420
DB 361 PALQNMKYDERQIARKRAIVSSMTSERENPDLLNPSRRRIAGSGNTFVEVNFKIDF 420

QY 421 NDMMKMKKQFTGGGKGGKGRNQNMQLKGM-----NLP 454
DB 421 NOAKQLMQGVMSG-----DMNKKMKQMGINPNLP 450
```

## RESULT 6

```
US-09-066-047-8
; Sequence 8, Application US/09066047A
; Patent No. 6306394
; GENERAL INFORMATION:
; APPLICANT: MORPHY, Cheryl
; STOREY, James
; BELTZ, Gerald A.
; COUGHLIN, Richard T.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF
; USE OF GRANULOCYTIC ERHLICHIA
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,047A
```

```
/
/ FILING DATE: 24-Apr-1998
/ CLASSIFICATION: <Unknown>
/
/ PRIORITY INFORMATION:
/ APPLICATION NUMBER: US 60/044,869
/ FILING DATE: 25-APR-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Superko, Colleen
/ REGISTRATION NUMBER: 39,850
/ REFERENCE/DOCKET NUMBER: 106,941,156
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 526-6000
/ TELEFAX: (617) 526-5000
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 420 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: <Unknown>
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-066-047-8

Query Match 37.2%; Score 846; DB 4; Length 420;
Best Local Similarity 40.8%; Pred. No. 6.3e-66;
Matches 173; Conservative 98; Mismatches 143; Indels 10; Gaps 3;

QY 3 FEGLSERLQATQMKRGKGLFEADIKIMREVRALFEADVNFYKVEFIKTVSERALG 62
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 2 FNSLTKGSSALQSLSKREISKDFDLVIDTQALDADVNLGVDFIENKSKIVG 61
QY 63 SDVMQSTPPGQVQIKIVODELTKMGENTSIINMSKPPVAVNVLGQAGKTTAGKLA 122
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 62 GDVKGVLPEQMWIKRIEELIENLSEKSLDLKGRIPAVIMVGLQGVGKTTNVKVA 121
QY 123 LLMRKYNKKPMLVAADIRPAAINQLOTGVKQIDIPVYSEGQVQKQIIVTNLKAKE 182
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 122 LRL-KKSKNPLVASLVYRPAERQLVLADGVGIDSLIVEEQKPLDIKRAMREARL 180
QY 183 EHLDSVITDAGRHLHIDEALMNEKVEIAKPEINMLVVDSTMGTQDAVNAESFDQDL 242
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 181 KGHVVLLDTAGRLHINQMDLCKVKEYSPEAVIIVLVDLSMGQDAVTVYKNEBELG 240
QY 243 VTGVTLTKLGDTRGGAAISRSVTQKPIKTVGMSKLDGLFLFHPERNASRIKMGDVL 302
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 241 ITGTTFRADGPRGGAISLMKIVACPIKFMSTGKPELDLDFYPDRIAREMLMGSDVA 300
QY 303 SLIEKAQDVQDEKAKDLEKMKRESSFTLDDFLQDQVKNLGLPDDIMKMPGK---- 357
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 301 SLVEKAVAGKDTINELQAKKGGKFDLDDIVQLKALNMGGINIMKFFIPAGNDIK 360
QY 358 NKMGLDLKNNSEKQIDHIKAIQSWTPAERNPDITLVNRSKRRIAGSGRSLSQVRLM 417
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 361 RKVAGI---ADDSKVDMTAIINSMTKQERANPEILNGARKARIAGAGVKVDVAVNALL 416
QY 418 KQFN 421
DB 417 KQFN 420

RESULT 7
US-08-317-401E-2
; Sequence 2, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
```

```
/
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10174-6401
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/317,401E
/ FILING DATE: 03-October-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Harrington, James J.
/ REGISTRATION NUMBER: 38,711
/ REFERENCE/DOCKET NUMBER: 4248,000-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212 867 0123
/ TELEFAX: 212 867 0298
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 534 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ ORIGINAL SOURCE:
/ ORGANISM: Aspergillus niger
/
US-08-317-401E-2

Query Match 25.8%; Score 586; DB 2; Length 534;
Best Local Similarity 33.6%; Pred. No. 4.6e-43;
Matches 155; Conservative 75; Mismatches 213; Indels 18; Gaps 9;

QY 1 MAPGLSERLQATQMKRGKGLFEAD- IKIMREVRALFEADVNFYKVEFIKTVSER 59
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 1 MVLQDLGRRINAAVNDLTNRNLDKQAFDDMIKEICAALLSADVNVELVQSRKSIKSS 60
QY 60 ALGSDVMQSTPPGQVQIKIVODELTKMGENTSIINMSKPPVAVNVLGQAGKTTAG 119
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 61 VNFASLPPAVNKKRLIQKAVFDELVSLVDPHAPFPKPKGRSNVIMFVGLQAGKTTTCT 120
QY 120 KLALLMKKYNK---KPMVLAADIYRPAAINQLOTGVKQIDIPVYSEGQVQKQIIVTNA 176
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 121 KLA---RHYQMGFTALVCACTFRAGAFDQKQATKAKIPYTGSLTQDPAIVAAEG 176
QY 177 LKHAKEEHLDFVIDTAGRLHIDEALMNEKVEIAKPEINMLVVDSTMGTQDAVNAES 236
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 177 VAKFKERFRIIIVDTSGRKHQEEELFTEMTQIQTAVTPQITLVLDDSTIGRAAEAGSSA 236
QY 237 FDDQLDVTGVTLLKDGDTRGGAALSRVSTQKPIKTVGMSKLDGLFLFHPERMASRI 296
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 237 FKATADFGAIIITKTGHAAGGGAISAVAATHPTIYLTGTGHELMDFRFPKAFIOKLL 296
QY 297 GNGDVLSLIEKAQOQ-DVDOEKAKDLEKMKRESSFTLDDFLQDQVKNLGLPDDIMKMP 355
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 297 GNGDMAGLVEHVQAVTKDSASAKETKHISEGTYTLDRPRENITSIMKAGPLSKLSMIP 356
QY 356 GN-NMKGLDLKNNSEKQIDHIKAIQSWTPAERNPDITLVN---SRKKRIAGSGRSLSQ 411
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 357 GLSNLTAGLDDDEGSMK-LRRMIYIFDSMTAAELDGDGKMFVEQPSRMVRIACSGGTVR 415
QY 412 EYVNLKMQFNDMKMMKMQFTGGGKGGKGRNQVQNMKGN 452
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 416 EYEDLLSQHRMAGMAKRV--GGQKKQMR--AQNMLKGN 452

RESULT 8
US-08-317-401E-4
; Sequence 4, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; APPLICANT: Yaver, Debbie Sue
```

;; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF  
;; TITLE OF INVENTION: ASPERGILLUS NIGER  
;; NUMBER OF SEQUENCES: 10  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: No. 59225610 No. 5922561disk of No. 5922561th America, Inc.  
;; STREET: 405 Lexington Avenue, Suite 6400  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 10174-6401  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/317,401E  
;; FILING DATE: 03-October-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Harrington, James J.  
;; REGISTRATION NUMBER: 38,711  
;; REFERENCE/DOCKET NUMBER: 4248,000-US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212 867 0123  
;; TELEFAX: 212 867 0298  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 552 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; ORIGINAL SOURCE:  
;; ORGANISM: Aspergillus niger  
;; US-08-317-401E-4

;; RESULT 9  
US-08-981-527A-8  
; Sequence 8, Application US/08981527A  
; Patent No. 6410262  
; GENERAL INFORMATION:  
; APPLICANT: Quax, Wilhelmus J.  
; APPLICANT: Kerkman, Richard  
; APPLICANT: Broekhuizen, Cornelis P.  
; TITLE OF INVENTION: No. 6410262el Secretion Factors for  
; TITLE OF INVENTION: Gram-Positive Microorganisms Genes Encoding Them and Methods  
; TITLE OF INVENTION: of using it  
; FILE REFERENCE: GCX322-US  
; CURRENT APPLICATION NUMBER: US/08/981,527A  
; CURRENT FILING DATE: 1997-12-16  
; PRIOR APPLICATION NUMBER: PCT/NL96/00278  
; PRIOR FILING DATE: 1996-07-05  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 329  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-08-981-527A-8  
  
Query Match 18.9%; Score 429.5; DB 4; Length 329;  
Best Local Similarity 34.0%; Pred. No. 1.1e-29;  
Matches 111; Conservative 64; Mismatches 116; Indels 35; Gaps 9;  
  
QY 4 EGLSERLQATMQRGKGLTEADIKIMREVRALFE-----ADNPKVVKYKEFI-- 53  
Db 16 DSVSEKFKDGLKTRNSFQNVYNDLVSRKVDDEFFEELEEVLSADVGTTVMELIDE 75  
  
QY 54 --KIVSERALGSDVMQSLTPGQGVKIVQDELKYL--MGCENTS--INMSKPPVYVMVG 108  
Db 76 LKKEVKKR-----NIODPKVKSVISEKLVEIYNSGDEQISELNIQDGRMLVILLG 127  
  
QY 109 LOGAGKTTTAGKLALLMRKKYKPKMLVAADIVRPAAINQLOTGVGKQIDIPVYSEGQVK 168  
Db 128 VNGVKTTTIGKLAHKM--KQEGKS SVLAAGDTFRAGALQLELVWGERGVPIKQTAGSD 186  
  
QY 169 PQQIVTNALKAHEHLDFVIIDTAGRLHIDEALMNEKVEKIAK-----PNEIMLVV 222  
Db 187 PAAVYDAVHAARADYLLICDTAGRLQNKVNLKMELEKVRVIEREVPAPHEVILLAL 246  
  
QY 223 DGMTQDAVNVASFDQDLDTGVTLTKLDGTRGGAALSIRSVTKPKIKFVGMSEKLDG 282  
Db 247 DATTCQNAQAQAKESKATNVGTGIALTKLDGTAKGGIVLAIRNELHIFPVKLVGLGKVD 306  
  
QY 283 LEFHPERMASRILGMDVLSLIERA 308  
Db 307 LOEFDPE---SYVYGL--FSDLVERA 327  
  
;; RESULT 10  
US-09-134-001C-4817  
; Sequence 4817, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GFC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4817  
; LENGTH: 415  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4817

Query Match 18.8%; Score 426.5; DB 4; Length 415;  
Best Local Similarity 31.4%; Pred. No. 2.8e-29;  
Matches 103; Conservative 76; Mismatches 122; Indels 27; Gaps 8;

QY 8 ERLOATMOKRGKGLTADIKIMREVRALFEADYVNFVKVEIKTVSERALGSDVMQ 67  
DB 100 QNFQEQNLNLIARYKRVDEDFEALDEEM---LITADVGFNTV---MTLEERMEAQRR 149  
QY 68 SLTPGQOVI-KIV-----QDELTKLGGENTSINMSKPPVVMVGLGAGKTTTAGKL 121  
DB 157 TEDLRVIVEKIVEIYHQDDYSEAMNIEDGLNVLIMVGVNGVGTITIGKL 209  
QY 122 ALLMRKYNKPMVAADYIPPAINQLQTVGKQIDIPYSEGQVQPOQIVTNALKHAK 181  
DB 210 AYRQOE-GKKVMLAAGDTFRAGAIQOLNVMGERVGVYSONEGSDPAAVYDA.NAAK 268  
QY 182 EHLDFVIIDTAGRLHIDEALMNLKEVKEIAK-----PNEIMLVVDSMTGQDAVNVAE 235  
DB 269 NKGVDIILICDTAGRLQNKSNLMQELDKMKRVINRAIPDAPHEALLCIDATTGQNALSQAR 328  
QY 236 SFDDQLDVGTVLTKDGTGGAALSIRSVTKPIKFKVMSKLDGLLELPHPERMASRI 295  
DB 329 SFKEVTNVSGIVLTIKDGTAKGGIVLAIRNELHIPVKYVGLGKMDLOLPENPE---SYV 385  
QY 296 LGM-GDVLSLTEKAQDVDOEK 322  
DB 386 YGLFADMIQNEIDPEISRNSSVESEE 413

## RESULT 11

US-09-007-476-2  
Sequence 2, Application US/09007476  
Patent No. 6159949

GENERAL INFORMATION:  
APPLICANT: Black, Michael T.  
TITLE OF INVENTION: No. 6159949el Ftsy  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Falk, Stephen T  
REGISTRATION NUMBER: 36,795  
REFERENCE/DOCKET NUMBER: GMI0079  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2488  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-09-007-476-2

Query Match 18.5%; Score 421; DB 4; Length 416;  
Best Local Similarity 31.4%; Pred. No. 8.4e-29;  
Matches 101; Conservative 78; Mismatches 121; Indels 22; Gaps 7;

QY 8 ERLOATMOKRGKGLTADIKIMREVRALFEADYVNFVKVEIKTVSERALGSDVMQ 67  
DB 97 QNFQEQNLNLIARYKRVDEDFEALDEEM---LITADVGFNTV---MTLEERMEAQRR 149  
QY 68 SLTPGQOVI-KIVQDELTKLGGE---NFSINMSKPPVVMVGLGAGKTTTAGKLALL 124  
DB 150 NIQWEDLRVIVEKIVEIYHQDDNSEAMNIEDGLNVLIMVGVNGVGTITIGKLA-Y 208  
QY 125 MRKYNKPMVAADYIPPAINQLQTVGKQIDIPYSEGQVQPOQIVTNALKHAKHEH 184  
DB 209 RYMEGKVKMLAAGDTFRAGAIQOLKRVGERVGVVDSQSEGSDPAAVMDAINAAKNG 269  
QY 185 LDFVIIDTAGRLHIDEALMNLKEVKEIAK-----PNEIMLVVDSMTGQDAVNVAE 238  
DB 269 VDILICDTAGRLQNKTNLMQELKEKRVKVINRAVPDAPHEALLCIDATTGQNALSQARNEK 328  
QY 239 DQDQDVGTVLTKDGTGGAALSIRSVTKPIKFKVMSKLDGLLELPHPER----MASR 294  
DB 329 EVTNVNGIVLTIKDGTAKGGIVLAIRNELHIPVKYVGLGKMDLOLPENPESYVYGLFAD 388  
QY 295 ILGMGDVLSLTEKAQDVDOEK 316  
DB 389 MIEQNDIEITVEN-DQIVTEEK 409

## RESULT 12

US-08-986-963-2  
Sequence 2, Application US/08986963  
Patent No. 5958730

GENERAL INFORMATION:  
APPLICANT: Rosteck Jr., Paul R.  
TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: US  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: December 8, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Webster, Thomas D.  
REGISTRATION NUMBER: 39,872  
REFERENCE/DOCKET NUMBER: X-11763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317/276-3334  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 425 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-986-963-2

Query Match 17.5%; Score 397.5; DB 2; Length 425;  
Best Local Similarity 32.5%; Pred. No. 9.8e-27;  
Matches 105; Conservative 64; Mismatches 109; Indels 45; Gaps 9;



[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```
Query Match          17.2% Score 390; DB 3; Length 277;
Best Local Similarity 33.5%; Pred. No. 2.4e-26;
Matches            92; Conservative    59; Mismatches   96; Indels      28; Gaps     5;
```

---

```
QY 48 VVKFKFIVTGERALGSDV-----MSTPQQVVIKIQVDLTKLM--- 87  
         :|::| | :|||  
db 1 VDEFFFLDELINDGVQVASNLTEELRYEAKLENKPKDALLRYYIEKLVELYEKD 60
```

Qy	86	GGEVTSINNENKDPVTVMWVGLOGAGKTTTACKLALLMRKKYKNKPMLVAAIDYRPAAIN	147
Db	61	SYDESIFHQDN-LATVMLFVGNGVGGKTSIGCLA-HRYKRTGKKYMLVAADTRACGAVD	118
Qy	148	OLQTVGQKIDIPYSEGQDVPPQOIVTNALKUAKEEHLDVVIDTGRULHDEALMNEKL	207
Db	119	QLAEWGHRVDVPVVTPPEKAPASVPFDGMERAVEGIDILMDTAGRLQNKDNLMAELE	178
Qy	208	EYKTIAK-----PNEIMLVWSDTSGDADVNAESFFDDQLDDVTGLTKLDGDTRFGAAL	261
Db	179	KIGRIIRRVPEAPHETFLADASTGQNLVQAQFEFKITPLGIVLKIDGTARGGWVL	238
Qy	262	SIRSVTGKPIKFVGMSEKLDGLELFHPERMASRIIL	296
Db	239	AIREELNIPVKLTIGFGEKIDDIGEFNSENFMKGLL	273

Search completed: February 21, 2003, 15:19:52  
Job time : 22 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: February 21, 2003, 15:19:07 ; Search time 13 seconds  
(without alignments)  
1087.444 Million cell updates/sec

Title: US-09-943-108a-2  
Perfect score: 2274  
Sequence: 1 MAFGLSERLQATMKRKG.....GKKGRNQNMKGNLPPF 455

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2261	99.4	455	10	US-09-815-242-13066 Sequence 13066, A
2	2235	98.3	455	10	US-09-815-242-5420 Sequence 5420, Ap
3	2235	98.3	455	10	US-09-815-242-12128 Sequence 12128, A
4	1418.5	62.4	472	10	US-09-815-242-4945 Sequence 4945, Ap
5	1418.3	62.4	477	10	US-09-815-242-10861 Sequence 10861, A
6	1307	57.5	523	10	US-09-815-242-13254 Sequence 13254, A
7	1306	57.4	466	10	US-09-815-242-13525 Sequence 13525, A
8	1297	57.0	523	10	US-09-814-041A-2 Sequence 2, Appli
9	1168	51.4	547	9	US-09-738-626-5759 Sequence 5759, Ap
10	1164.5	51.2	453	10	US-09-815-242-10263 Sequence 10263, A
11	1150	50.6	457	10	US-09-815-242-11948 Sequence 11948, A
12	1138	50.0	462	10	US-09-815-242-10965 Sequence 10965, A
13	1127.5	49.6	501	10	US-09-815-242-14089 Sequence 14089, A
14	854	37.6	448	10	US-09-815-242-11562 Sequence 11562, A
15	432	19.0	497	10	US-09-815-242-10378 Sequence 10378, A
16	418	18.4	416	10	US-09-815-242-12126 Sequence 12126, A
17	416	18.3	316	10	US-09-815-242-5299 Sequence 5299, Ap
18	404.5	17.8	442	10	US-09-815-242-10904 Sequence 10904, A
19	404	17.8	455	10	US-09-815-242-11781 Sequence 11781, A

20	403.5	17.7	313	10	US-09-815-242-4992 Sequence 4992, Ap
21	397.5	17.5	425	10	US-09-815-242-13263 Sequence 13263, A
22	397.5	17.5	430	10	US-09-827-663-2 Sequence 2, Appli
23	396.5	17.4	429	10	US-09-815-242-13589 Sequence 13589, A
24	395.5	17.4	414	10	US-09-815-242-11094 Sequence 11094, A
25	390	17.2	277	10	US-09-827-663-4 Sequence 4, Appli
26	376	16.5	510	9	US-09-738-626-5759 Sequence 5759, Ap
27	151	6.6	1945	9	US-09-927-597-2 Sequence 2, Appli
28	151	6.6	1979	9	US-09-927-597-4 Sequence 4, Appli
29	140.5	6.2	2025	10	US-09-815-242-5703 Sequence 5703, Ap
30	140.5	6.2	3158	10	US-09-815-242-12611 Sequence 12611, A
31	134	5.9	705	10	US-09-815-242-5334 Sequence 5334, Ap
32	134	5.9	713	10	US-09-815-242-12322 Sequence 12322, A
33	132	5.8	856	10	US-09-815-242-11310 Sequence 11310, A
34	130.5	5.7	1786	9	US-09-742-096-3 Sequence 3, Appli
35	129.5	5.7	2437	10	US-09-815-242-5834 Sequence 5834, Ap
36	129.5	5.7	6281	10	US-09-815-242-12996 Sequence 12996, A
37	129	5.7	856	10	US-09-815-242-11489 Sequence 11489, A
38	127.5	5.6	866	12	US-10-007-693-113 Sequence 113, App
39	126	5.5	1342	10	US-09-815-242-10438 Sequence 10438, A
40	124.5	5.5	800	10	US-09-815-242-5349 Sequence 5349, Ap
41	124.5	5.5	800	10	US-09-815-242-12139 Sequence 12139, A
42	124.5	5.5	800	10	US-09-815-242-13136 Sequence 13136, A
43	124.5	5.5	1212	10	US-09-815-242-5279 Sequence 5279, Ap
44	124.5	5.5	1482	10	US-09-815-242-12484 Sequence 12484, A
45	121.5	5.3	592	10	US-09-861-451A-72 Sequence 72, Appli

ALIGNMENTS

RESULT 1  
US-09-815-242-13066  
; Sequence 13066, Application US/09815242  
; Patent No. US20020061569A1

; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.01A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13066

; LENGTH: 455  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-13066

Query Match 99.4%; Score 2261; DB 10; Length 455;  
Best Local Similarity 99.3%; Pred. No. 1.3e-154;



Matches	452;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;
QY	1	MAFEGLSERLQATQMKWRGKGLTEADIKIMREVR	LALFEADVNFKVVKEFIKTVSERA	60					
Db	1	MAFEGLSERLQATQMKWRGKGLTEADIKIMREVR	LALLEADVNFKVVKEFIKTVSERA	60					
QY	61	LGSVMSLSLPGQGVTKIVQDELTKLGGENTSI	NMSNPPTVVMWGLQAGKTTTAK	120					
Db	61	LGSVMSLSLPGQGVTKIVQDELTKLGGENTSI	NMSNPPTVVMWGLQAGKTTTAK	120					
QY	121	LALLMRKYNKPKMLVAADIYPPAAINQLQ	TVGQKIDIPVYSEGQVQKPOIVTNALKHA	180					
Db	121	LALLMRKYNKPKMLVAADIYPPAAINQLQ	TVGQKIDIPVYSEGQVQKPOIVTNALKHA	180					
QY	161	KEEHLDFVIIIDTAGRLIHIDEALMNELKEV	KELAKPNEIMLVVDSMTGQADVNAESFDQ	240					
Db	161	KEEHLDFVIIIDTAGRLIHIDEALMNELKEV	KDIAKPNEIMLVVDSMTGQADVNAESFDQ	240					
QY	241	LDVTGVTLTKLDGDTGGGAALSISVTQKPK	IKFVGMSEKLDGLGFHPERMAKRIJLGMG	300					
Db	241	LDVTGVTLTKLDGDTGGGAALSISVTQKPK	IKFVGMSEKLDGLGFHPERMAKRIJLGMG	300					
QY	301	VLSLIEKAQDVQDEKAKDLEKKMRESSFT	LDLDFLEQLDQVKNLGLDDIMKIPGMNKK	360					
Db	301	VLSLIEKAQDVQDEKAKDLEKKMRESSFT	LDLDFLEQLDQVKNLGLDDIMKIPGMNKK	360					
QY	361	KGDLKANSEKQIDHILKALIQSMTPAERN	PDPTLVNSRKKRIKAGSGESLOEVNRLMKOF	420					
Db	361	KGDLKANSEKQIDHILKALIQSMTPAERN	PDPTLVNSRKKRIKAGSGESLOEVNRLMKOF	420					
QY	421	NDMKMMKMQFTGGGKGKKRNONMNLKGMN	LPF	455					
Db	421	NDMKMMKMQFTGGGKGKKRNONMNLKGMN	LPF	455					

```

RESULT 2
US-09-815-242-5420
; Sequence 5420, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,737
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5420
; LENGTH: 455
; TYPE: PRF
; ORGANISM: Staphylococcus aureus

```

US-09-815-242-5420

```
Query Match          98.3%; Score 2235; DB 10; Length 455;
Best Local Similarity 98.2%; Pred. NO. 9.3e-153;
Matches 447; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

QY	1	MAFEGLSERLQATQMKRGKGLTEADIKIMREVRVRLALFEADVNFVKVYKEIKTVSERA	60
Db	1	NAFEGLSERLQATQMKRGKGLTEADIKIMREVRVRLALFEADVNFVKVYKEIKTVSERA	60
QY	61	LGSDWQSLTPGQOVKIVQDELTKLMGGENTSINMSNKPTPTVMVGLQGAGKTTAGK	120
Db	61	LGSDWQSLTPGQOVKIVQDELTKLMGGENTSINMSNKPTPTVMVGLQGAGKTTAGK	120
QY	121	LALLMRKYNKPKMLVAADIYRPAAINQLQTVGKQID:PVYSEGQDVKPOQIVTNALKHA	180
Db	121	LALLMRKYNKPKMLVAADIYRPAAINQLQTVGKQID:PVYSEGQDVKPOQIVTNALKHA	180
QY	181	KEEHLDEVIIDTAGRLHIDEALMNEKEVKEIAKENEIMLVDSMTGQDYNVAESDDQ	240
Db	181	KEEHLDEVIIDTAGRLHIDEALMNEKEVKEIAKENEIMLVDSMTGQDYNVAESDDQ	240
QY	241	LDVTGVTILKLDGTRGGAALSIKRSVTQKPIKFVGWSEKLDGLLELPHPERMASRILGMD	300
Db	241	LDVTGVTILKLDGTRGGAALSIKRSVTQKPIKFVGWSEKLDGLLELPHPERMASRILGMD	300
QY	301	VLSLIEKAQODVDQKADLEKKMRSSFTLDDFLEQLDQVKNLGPIDDIKMKMTPGNMKM	360
Db	301	VSRLEIEKAQODVDQKADLEKKMRSSFTLDDFLEQLDQVKNLGPIDDIKMKMTPGNMKM	360
QY	361	KGDLKLNMSKQIDHIIKAIIOQMTPAERNPPTLVNSKKRIAKGSGRSLOEVNRLMKQF	420
Db	361	KGDLKLNMSKQIDHIIKAIIOQMTPAERNPPTLVNSKKRIAKGSGRSLOEVNRLMKQF	420
QY	421	NDMKKMKQFTGGGKKGKGRNQOMNLMKGMNLPF	455
Db	421	NDMKKMKQFTGGGKKGKGRNQOMNLMKGMNLPF	455

### RESULT 3

```

US-09-815-242-12128
; Sequence 12128, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Irwack, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191.078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206.848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207.727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242.578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253.625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257.931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269.308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSQ for Windows Version 4.0

```

; SEQ ID NO 12128

; LENGTH: 455

; TYPE: PRF

; ORGANISM: Staphylococcus aureus

US-09-815-242-12128

Query Match

Best Local Similarity 98.3%; Score 2235; DB 10; Length 455;

Matches 447; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAPEGLSERLOATQMKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60  
Db 1 MAPEGLSERLOATQMKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60  
QY 61 LGSQVMSLTGQOVIVQDELTKLMGGENTISNMSNKPPTVVMVGLQGAGKTTTAGK 120  
Db 61 LGSQVMSLTGQOVIVQDELTKLMGGENTISNMSNKPPTVVMVGLQGAGKTTTAGK 120  
QY 121 LALLMRKYNKPMVAADYIRPAAINQLOTGVKQIDIPVYSGDQVQKQOIVTNALKHA 180  
Db 121 LALLMRKYNKPMVAADYIRPAAINQLOTGVKQIDIPVYSGDQVQKQOIVTNALKHA 180  
QY 181 KEHLDVFIIDTAGRLHIDEALMDELKEVKEIAKPNELMVLVDSMTGQAVNVAESFDDQ 240  
Db 181 KEHLDVFIIDTAGRLHIDEALMDELKEVKEIAKPNELMVLVDSMTGQAVNVAESFDDQ 240  
QY 241 LDVGTGVTLLKLDGTRGGAALSIRSVTKQPIKFGVMSKLDGLLELPHPERMASRIILGMD 300  
Db 241 LDVGTGVTLLKLDGTRGGAALSIRSVTKQPIKFGVMSKLDGLLELPHPERMASRIILGMD 300  
QY 301 VLSLEKAQDVQDEKAKDEKEMRESSFTLDDLEQLDQVKNLGPDDIMKMPGNMKM 360  
Db 301 VLSLEKAQDVQDEKAKDEKEMRESSFTLDDLEQLDQVKNLGPDDIMKMPGNMKM 360  
QY 361 KGLDLKLNSEKQIDHIKAIQSMTPAERNPDTLNVSRKKRIAKSGRSLSQEVNRLMKQF 420  
Db 361 KGLDLKLNSEKQIDHIKAIQSMTPAERNPDTLNVSRKKRIAKSGRSLSQEVNRLMKQF 420  
QY 421 NDMMKMKMKQFTGGGKGGKGRNQMNKLGMLNLPF 455  
Db 421 NDMMKMKMKQFTGGGKGGKGRNQMNKLGMLNLPF 455

RESULT 4

US-09-815-242-4945

; Sequence 4945; Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4945

; LENGTH: 472

; TYPE: PRF

; ORGANISM: Enterococcus faecalis

US-09-815-242-4945

Query Match

Best Local Similarity 59.0%; Pred. No. 38-94;

Matches 273; Conservative 86; Mismatches 89; Indels 15; Gaps 2;

QY 1 MAPEGLSERLOATQMKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60  
Db 1 MAPEGLSERLOATQMKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60  
QY 61 LGSQVMSLTGQOVIVQDELTKLMGGENTISNMSNKPPTVVMVGLQGAGKTTTAGK 120  
Db 61 LGSQVMSLTGQOVIVQDELTKLMGGENTISNMSNKPPTVVMVGLQGAGKTTTAGK 120  
QY 121 LALLMRKYNKPMVAADYIRPAAINQLOTGVKQIDIPVYSGDQVQKQOIVTNALKHA 180  
Db 121 LALLMRKYNKPMVAADYIRPAAINQLOTGVKQIDIPVYSGDQVQKQOIVTNALKHA 180  
QY 181 KEHLDVFIIDTAGRLHIDEALMDELKEVKEIAKPNELMVLVDSMTGQAVNVAESFDDQ 240  
Db 181 KEHLDVFIIDTAGRLHIDEALMDELKEVKEIAKPNELMVLVDSMTGQAVNVAESFDDQ 240  
QY 241 LDVGTGVTLLKLDGTRGGAALSIRSVTKQPIKFGVMSKLDGLLELPHPERMASRIILGMD 300  
Db 241 LDVGTGVTLLKLDGTRGGAALSIRSVTKQPIKFGVMSKLDGLLELPHPERMASRIILGMD 300  
QY 301 VLSLEKAQDVQDEKAKDEKEMRESSFTLDDLEQLDQVKNLGPDDIMKMPGNMKM 360  
Db 301 VLSLEKAQDVQDEKAKDEKEMRESSFTLDDLEQLDQVKNLGPDDIMKMPGNMKM 360  
QY 361 KGLDLKLNSEKQIDHIKAIQSMTPAERNPDTLNVSRKKRIAKSGRSLSQEVNRLMKQF 420  
Db 361 KGLDLKLNSEKQIDHIKAIQSMTPAERNPDTLNVSRKKRIAKSGRSLSQEVNRLMKQF 420  
QY 421 NDMMKMKMKQFTGGGKGGKGRNQMNKLGMLNLPF 459  
Db 421 NDMMKMKMKQFTGGGKGGKGRNQMNKLGMLNLPF 459

RESULT 5

US-09-815-242-10861

; Sequence 10861; Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10861
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-815-242-10861

Query Match 62.4%; Score 1418.5; DB 10; Length 477;
Best Local Similarity 59.08; Pred. No. 3e-94; Mismatches 89; Indels 15; Gaps 2;
Matches 273; Conservative 86;

QY 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVEALFADVNFKVKEFKTVSERA 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1 MAFESLTNRLOQAMSKIRKKGKSEADVKEEMREIRLALLEADVNLOVVKDFIKVRERA 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 61 LGSVMSLTPGQOVTKIVQDELTKMGCGTNTSINMKNKPPVTVMVVGLOGAGKTTTACK 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 61 VGVLESLSPAQIVKIVDELTGSETVELNKSPIPTVIMTGLGAGKTTTFTGK 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 121 LALIMRKYKPKMELVAADIYRPAAINQLQTVGKQIDIPVYSEGQVQKPOQIVTNALKHA 180
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 121 LAKHLAKTENAPLLIAGDVYRPAIDQLKVLQQLLEVYFDMGTNDANVEIVRQGLALA 180
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 181 KEELDFVIDTAGRLHIDEALMNELKEVKEIAKNEIMLVVDSMTGQDANVAESFDDQ 240
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 181 KKKNDYVIDTAGRLHIDEALMDELQKIELANPNEILLVVDAMTGDANVADSFNQ 240
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 241 LDVTGVTLLKLDGTRGGAALSIRSVTKPIKFGVSGSEKLDGLLELHPERMASRIILMGD 300
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 241 LGITGVVTKLDGTRGGAALSIRAVTGAPIKFGVSGSEKLDGLLELHPDRMSRIILMGD 300
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 301 VLSLIEKAQDVQDKAKDLEKKMRSSFTLDDFLEQLDQVKNLGLDIDIMKMPGNMK 360
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 301 MLTIEKAQDVDEKKAELAQMKENSDFDNDFIEQLDQVNGMGPIDELDKMIPGNM 360
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 361 KGLDKLNMSEKQIDHKAIIOSMTPAERNPDPLNVSRKRIAKSGESLSQEVNRLMKQF 420
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 361 PGIENTVKVPDKVARKRAVLMSPTAERNPDPLNPSRRRIAAAGSGSVNVEVRMIKQF 420
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 421 NDMKKMMKQFT-----GGGKGGKGRNQNMNLK 449
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 421 KESKMMQMSKGMNIFGMQMLGGGVKGLGKM-AMNRMK 462
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 6
US-09-815-242-13254
; Sequence 13254, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13254
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-815-242-13254

Query Match 57.5%; Score 1307; DB 10; Length 523;
Best Local Similarity 55.1%; Pred. No. 3.3e-86;
Matches 253; Conservative 90; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVEALFADVNFKVKEFKTVSERA 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1 MAFESLTNRLOQAMSKIRKKGKISESDVQATKEIRLALLEADVALFVVKDFIKVRERA 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 61 LGSVMSLTPGQOVTKIVQDELTKMGCGTNTSINMKNKPPVTVMVVGLOGAGKTTTACK 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 61 VGVHEVIDTLNPAQIILKIVDELTAVLGSDTAELIKSPKPTIIMVVGLOGAGKTTTACK 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 121 LALIMRKYKPKMELVAADIYRPAAINQLQTVGKQIDIPVYSEGQVQKPOQIVTNALKHA 180
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 121 LANKLKEENARPLVAADIYRPAIDQLKTLGQQIDVPFALGTEVPAVEIVRQGLEQA 180
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 181 KEELDFVIDTAGRLHIDEALMNELKEVKEIAKNEIMLVVDSMTGQDANVAESFDDQ 240
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 181 QTNHEVIDTLNPAQIILKIVDELTAVLGSDTAELIKSPKPTIIMVVGLOGAGKTTTACK 240
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 241 LDVTGVTLLKLDGTRGGAALSIRSVTKPIKFGVSGSEKLDGLLELHPERMASRIILMGD 300
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 241 LEVTGVVTKLDGTRGGAALSIRVHITGPKETGTGKIDTIDTFHPDRMSRIILMGD 300
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 301 VLSLIEKAQDVQDKAKDLEKKMRSSFTLDDFLEQLDQVKNLGLDIDIMKMPGNMK 360
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 301 MLTIEKAQDVDEKKAELMAEKRENTDFDNDFIDQLDQVNGMGPIDELDKMIPGNM 360
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 361 KGLDKLNMSEKQIDHKAIIOSMTPAERNPDPLNVSRKRIAKSGESLSQEVNRLMKQF 420
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 361 PALQNMKVDERQLARKRAIVSSNTPPEERNPDPLNPSRRRIAAAGSGNTFVFNKFKIDF 420
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 421 NDMKKMMKQFTGGGKGGKGRNQNMNLKGM-----NLP 454
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 421 NOAKQLMQGVMS-----DNKMMKMGGINPNLNP 450
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 7
US-09-815-242-13525
; Sequence 13525, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13525
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13525
```

```
Query Match 57.4%; Score 1306; DB 10; Length 466;
Best Local Similarity 54.9%; Pred. No. 3.4e-86;
Matches 252; Conservative 91; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGLSERLOATMOKMRGKGLTADIKIMREVRALFEADYNFKVKEFIKTVSERA 60
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 1 MAFESLTERLQNVFKLRKKGKISESDVOEATKEIRLLEADVALPVVKDFIKKVRERA 60
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 61 LGSVMQSLTPGQOVKIVQDELTKLGGENTSIKMSKPKPTVVMVGLQAGKTTAGK 120
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 61 VGEHVIDTLNPAQIILKIVDELTAVLGSDTAEIKSPKIPITIMVGLQAGKTTAGK 120
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 121 LALLMRKKNKPMVAADYIRPAINQLQTVGKQIDIPYSGDQVKPQOIVTNALKHA 180
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 121 LANKLKEENARPLMAADIYRPAADQLKTLGQOIDVPFALGTEYPAVEIVRQGLEA 180
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 181 KEELHDFVIDTAGRLHIDEALMNEKVEKIAKPNEMLVDSMTGODAVNVAESFDDQ 240
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 181 QTNHNDYVLIDTAGRLQIDELMNEKVEKIAKPNEMLVDSMTGODAVNVAESFDDQ 240
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 241 LDVTGVTLKIDGTGGAALSRVTKPIKFGVMSKLDGLFHPERMASRILGMGD 300
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 241 LEVTVGVTILKIDGTGGAALSRVTKPIKFGVMSKLDGLFHPERMASRILGMGD 300
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 301 VLSLIEKAQDVQEKAKDEKKKRESSTLDDFLQDQVKNLGLPDDIMKMPGMANN 360
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 301 MTLTIEKASQYDQKALEMAEKRENTFDFNDIDQDQVKNLGLPDDIMKMPGMANN 360
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 361 KGLDLNMSKQIDHIAKIIQSMTPAERNPDTLNVSRKKRIKAGSGRSLOEVNRLMKQF 420
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 361 PALQNMKVDQRQIARKRAIVSSMTSEERENPDLLNPSRRRIIAAGSGNFTVEVVKIKDF 420
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 421 NDMMKMMKQFTGGGKGGKKNQNMNLKGM-----NLP 454
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 421 NOAKQLMQGVMSG-----DMNKKMKQMGINPNLNP 450
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
```

```
RESULT 8
US-09-814-041A-2
; Sequence 2, Application US/09814041A
; Patent No. US20020103104A1
; GENERAL INFORMATION:
; APPLICANT: CHEEVER, CHRISTY
; APPLICANT: PECTEAU, DOUGLAS A.
; APPLICANT: LI, HU
; APPLICANT: PAYNE, DAVID J.
; APPLICANT: STEEL, ANGELA
; APPLICANT: WANG, LEI
; TITLE OF INVENTION: METHODS USING THE SRP POLYNUCLEOTIDES
; AND POLYPEPTIDES COMPOUNDS MODULATING THEIR ACTIVITY
; FILE REFERENCE: GM50069
```

```
; CURRENT APPLICATION NUMBER: US/09/814,041A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/191,008
; PRIOR FILING DATE: 2000-03-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-814-041A-2

Query Match 57.0%; Score 1297; DB 10; Length 523;
Best Local Similarity 54.7%; Pred. No. 1.7e-85;
Matches 251; Conservative 91; Mismatches 103; Indels 14; Gaps 2;

QY 1 MAFEGLSERLOATMOKMRGKGLTADIKIMREVRALFEADYNFKVKEFIKTVSERA 60
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 1 MAFESLTERLQNVFKLRKKGKISESDVOEATKEIRLLEADVALPVVKDFIKKVRERA 60
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 61 LGSVMQSLTPGQOVKIVQDELTKLGGENTSIKMSKPKPTVVMVGLQAGKTTAGK 120
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 61 VGEHVIDTLNPAQIILKIVDELTAVLGSDTAEIKSPKIPITIMVGLQAGKTTAGK 120
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 121 LALLMRKKNKPMVAADYIRPAINQLQTVGKQIDIPYSGDQVKPQOIVTNALKHA 180
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 121 LANKLKEENARPLMAADIYRPAADQLKTLGQOIDVPFALGTEYPAVEIVRQGLEA 180
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 181 KEELHDFVIDTAGRLHIDEALMNEKVEKIAKPNEMLVDSMTGODAVNVAESFDDQ 240
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 181 QTNHNDYVLIDTAGRLQIDELMNEKVEKIAKPNEMLVDSMTGODAVNVAESFDDQ 240
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 241 LDVTGVTLKIDGTGGAALSRVTKPIKFGVMSKLDGLFHPERMASRILGMGD 300
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 241 LEVTVGVTILKIDGTGGAALSRVTKPIKFGVMSKLDGLFHPERMASRILGMGD 300
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 301 VLSLIEKAQDVQEKAKDEKKKRESSTLDDFLQDQVKNLGLPDDIMKMPGMANN 360
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 301 MTLTIEKASQYDQKALEMAEKRENTFDFNDIDQDQVKNLGLPDDIMKMPGMANN 360
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 361 KGLDLNMSKQIDHIAKIIQSMTPAERNPDTLNVSRKKRIKAGSGRSLOEVNRLMKQF 420
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 361 PALQNMKVDQRQIARKRAIVSSMTSEERENPDLLNPSRRRIIAAGSGNFTVEVVKIKDF 420
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 421 NDMMKMMKQFTGGGKGGKKNQNMNLKGM-----NLP 454
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 421 NOAKQLMQGVMSG-----DMNKKMKQMGINPNLNP 450
||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
```

```
RESULT 9
US-09-738-626-5755
; Sequence 5755, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKURO
; APPLICANT: OCHIALI, KEIJO
; APPLICANT: YOKOL, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
```

; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent In ver. 3.0  
; SEQ ID NO 5755  
; LENGTH: 547  
; TYPE: PRF  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5755

Query Match 51.4%; Score 1168; DB 9; Length 547;  
Best Local Similarity 49.1%; Pred. No. 3.2e-76;  
Matches 237; Conservative 85; Mismatches 125; Indels 36; Gaps 5;

Qy 3 FEGLSERLOATMOKRGKGLTEADIKIMREVRRLALFEADVNFVKVKEFIKTVSERALG 62  
Db 2 FESLSRLNSALSGRGKGLTEADINATREIRLALLEADVSLTVYRAFINKERAVG 61  
Qy 63 SDYMSQSLTPGQOVIIVDELTKIMGENTSIINMSKPPPTVMMVGLQGAGKTTAGKLA 122  
Db 62 AEYSQALNPAQOQVIKIVNEELVQILGETTRJSLAKNPPTVINLAGLQAGKTTAGKLS 121  
Qy 123 LLMRRKYNKPMVAADYTPAAINOLQTVGKOIDIPVY-----SEGD 165  
Db 122 KHLVKQ-GHTPMLVACDLQRGAQQLOIVGERAGVTTFADPDGTSIDSLEHEKGTSHGD 180  
Qy 166 QVQPOIVTNALKHAKHEHLDVLIIDTAGRLHIDBALMNEIKVEIKAPNEIMLVDSM 225  
Db 181 ---PVEVARAGTEEAARTQHDIVIVDTAGRLGIDETLMTQARTIREAINPDEVLFVIDSM 237  
Qy 226 TGDADVNAESDDLDVTVTLTKLDGTRGGAALSIRSVTKPIKVEVMSSEKLDGLEL 285  
Db 238 IGDADVTAEPGROGVDFTGVTLTKLDGDAAGGAALSIREVTGKPIINFAGSTGEKLDDEPV 297  
Qy 286 FHPERMASRLGMDVSLIEKAQODVDEKADLEKKMRESSTLDDFLQDLDQVKNLG 345  
Db 298 FHPERMASRLGMDVSLIEQAENVMDQEAQKLSGSELTLEDFDQMLIRMG 357  
Qy 346 PLDDIMKMTPGNKKGLDKINMSEKOIDHIKAIQSMTPAERNPDTLVNRSKKRIAG 405  
Db 358 PIGNILKMLPGGKQMSQADN-VDEKQDLRIQAILTGMTPAERNPKILNASRRKRIAG 416  
Qy 406 SGRSLOEVNRLMKQFNDKMKMQFTG-----GGKGGKGGKKNQNMKGM 451  
Db 417 SGVTSVANKVJVERFEARKVMQMAQGFQMGFGSRSATKQKQAGKNGKPKAKGP 476  
Qy 452 NLP 454  
Db 477 TQP 479

## RESULT 10

US-09-815-242-10263  
; Sequence 10263, Application US/09815242  
; Patent No. US20020061569A1

## GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELTRA 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10263  
; LENGTH: 453  
; TYPE: PRF  
; ORGANISM: Escherichia coli  
US-09-815-242-10263

Query Match 51.2%; Score 1164.5; DB 10; Length 453;  
Best Local Similarity 51.0%; Pred. No. 4.4e-76;  
Matches 231; Conservative 94; Mismatches 121; Indels 7; Gaps 2;

Qy 3 FEGLSERLOATMOKRGKGLTEADIKIMREVRRLALFEADVNFVKVKEFIKTVSERALG 62  
Db 2 FDLNLTDLRLSRTURNISGRGLTDEDNVKTDLREVRMLLEADVLPVVVREFINRVKEKAVG 61  
Qy 63 SDYMSQSLTPGQOVIIVDELTKIMGENTSIINMSKPPPTVMMVGLQGAGKTTAGKLA 122  
Db 62 HEVYNKSLTPGQOEFYKIVNELVAAMGEENQTLNLAQPPAVTLNAGLOGAGKTTSGKLG 121  
Qy 123 LLMRRKYNKPMVAADYTPAAINOLQTVGKOIDIPVYSEGQVKKPQVTVNALKHAK 182  
Db 122 KFLREXKHKVIVSADYTPAAIKOLETLAEQVGVDFPDSVGVQKPDVTVNAALKEAKL 181  
Qy 183 EHLDFVLIIDTAGRLHIDBALMNEIKVEIKAPNEIMLVDSMTGODAVNVAESDDQLD 242  
Db 182 KFDVULVLDITAGRLHIDBALMNEIKVEIKAPNEIMLVDSMTGODAVNVAESDDQLD 241  
Qy 243 VTGVTTLTKLDGTRGGAALSIRSVTKPIKVEVMSSEKLDGLELPHPERMASRLGMDV 302  
Db 242 LTGVLTVDGDRGGAALSIRHITGKPIKFLGVGEKTEALEPDPDRIASRLGMDV 301  
Qy 303 SLIEKAQODVDEKADLEKKMRE-SSFTLDDFLQDLDQVKNLGFLDDIMKMTPGNKKM 361  
Db 302 SLIEDTESVDRAQAEKLSAKKKGDFDLNDFLQLRQMKNMGMASLAKGLPGMGQIP 361  
Qy 362 GLDKLNKSEKOIDHIKAIQSMTPAERNPDTLVNRSKKRIAGSGRSLOEVNRLMKQFN 421  
Db 362 DNVKSQDDKVLVRMEALINSMTWKERAPEIKGSRRRRIAGCGMQGVQDVNRLKQFD 421  
Qy 422 DMKKMMKQFTGGGKGGKKNQNMKGMNLP 454  
Db 422 DMQRMK-----KMKGGMAKMRSMKGMMP 448

## RESULT 11

US-09-815-242-11948  
; Sequence 11948, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELTRA 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11948
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11948

```

```

Query Match          50.6%; Score 1150; DB 10; Length 457;
Best Local Similarity 50.5%; Pred. No. 4.9e-75;
Matches 231; Conservative 93; Mismatches 121; Indels 12; Gaps 3;

```

```

QY 3 FEGLSERLOAQYQMRGKGLTEADIKIMREVRALFEADVNFVKVKEFKITVSRALG 62
DB 2 FENLDRLSQTLRHVTQKAKLTEDNKTDLREVFMALLEADVLPVYKDFNKVRAVG 61
QY 63 SDVMQSLTPGQOQVIKIYQDELTKLMGGENTSINMSKPPVVMVMMVLOGAGKTTAGKLA 122
DB 62 TEVSLSLTPGQAFYKIVRAEELMGAANEDLALSAPPVAVILMAGLOGAGKTTAGKLA 121
QY 123 LMRKXKPKMLVAADYRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 182
DB 122 RFLKERRKKSVMYVADYRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 181
QY 183 EHLDFVITDAGRLHIDEALMNELEKVEIKAEKPNELMVLVDSMTGQDANVVAESFDOLD 242
DB 182 KFDVIVVITDAGRLHIDEALMNELEKVEIKAEKPNELMVLVDSMTGQDANVVAESFDOLD 241
QY 243 VTGVTITKLDGTRGGAALSIRSVTKPIKFGVMSKLDGLLELPHPERMASRILGMDVL 302
DB 242 LTGVVLTQVGDARGGAALSIRSVTKPIKFLGMSKSEALDPFHDPDRVASRILGMDVL 301
QY 303 SLIEAQODVDQEKAKDLEKKMRE--SSFTLDDFLEQLDQVKNLGPLDDITMKIP---GMN 358
DB 302 SLIEAQONLDRKAEKLAKKIKKGGFLEDFRDLQOQMKMGGLGGMLDKPLMLGSEVN 361
QY 359 KMKGLDKLNMSEKQIDHIKAIQSMPTFAERNPNPTLNVSRKKRIAGSGRSLOEYNRLMK 418
DB 362 LAQMGAQAAGAEKQFKQWEALINSMTPGERRDPPEMISGRKRRIALGSGTQVDVGRLIK 421
QY 419 QFNDMKKMKMFTGGGKGGKGGKGNQNMKMLP 455
DB 422 QHKQMKMKMKVYTAG-----GNMKMREGMSNF 450

```

```

RESULT 12
US-09-815-242-10965
; Sequence 10965, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes

```

```

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10965
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10965

```

```

Query Match          50.0%; Score 1138; DB 10; Length 462;
Best Local Similarity 50.0%; Pred. No. 3.6e-74;
Matches 225; Conservative 93; Mismatches 122; Indels 10; Gaps 2;

```

```

QY 3 FEGLSERLOAQYQMRGKGLTEADIKIMREVRALFEADVNFVKVKEFKITVSRALG 62
DB 2 FENLDRLSQTLRHVTQKAKLTEDNKTDLREVFMALLEADVLPVYKDFNKVRAVG 61
QY 63 SDVMQSLTPGQOQVIKIYQDELTKLMGGENTSINMSKPPVVMVMMVLOGAGKTTAGKLA 122
DB 62 EVNKSLLTPQCEFLKIVQRELEKAMGEANESLNATQPPAVILMAGLOGAGKTTAGKLA 121
QY 123 LMRKXKPKMLVAADYRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 182
DB 122 KFLERKXKPKMLVAADYRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 181
QY 183 EHLDFVITDAGRLHIDEALMNELEKVEIKAEKPNELMVLVDSMTGQDANVVAESFDOLD 242
DB 182 KFDVIVVITDAGRLHIDEALMNELEKVEIKAEKPNELMVLVDSMTGQDANVVAESFDOLD 241
QY 243 VTGVTITKLDGTRGGAALSIRSVTKPIKFGVMSKLDGLLELPHPERMASRILGMDVL 302
DB 242 LTGVVLTQVGDARGGAALSIRSVTKPIKFLGMSKSEALDPFHDPDRVASRILGMDVL 301
QY 303 SLIEAQODVDQEKAKDLEKKMRE--SSFTLDDFLEQLDQVKNLGPLDDITMKIPGMNKK 361
DB 302 SLIEDLERSVDREKAEKNAQKFKKGGDFTLDDFLEQLDQVKNLGPLDDITMKIPGMNKK 361
QY 362 GLDKLNMSEKQIDHIKAIQSMPTFAERNPNPTLNVSRKKRIAGSGRSLOEYNRLMKQFN 421
DB 362 EHVKNQVDDKAEKNAQKFKKGGDFTLDDFLEQLDQVKNLGPLDDITMKIPGMNKKQFN 421
QY 422 DMKKMKMFTGGGKGGKGGKGNQNMKMLG 451
DB 422 EMQRMKMKKGG-----MAKMRGM 442

```

```

RESULT 13
US-09-815-242-14089
; Sequence 14089, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.

```



```
RESULT 15
US-09-815-242-10378
; Sequence 10378, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITFA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-08-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10378
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10378

Query Match          19.0%; Score 432; DB 10; Length 497;
Best Local Similarity 31.3%; Pred. No. 1.5e-23;
Matches 98; Conservative 72; Mismatches 119; Indels 24; Gaps 5;

QY  4  EGLSERLOATMQRG-----KGLTEADIKIMREYRLALFADVNFKVKEFI 53
Db  193  EGFFARLRSLTKENLGSFISLFRGKKIDDD---LFELEEQLLIADVGVETTRKII 249

QY  54  KTVSRALGSDVMOSTPFGQVIVKIVODELTKLGGENTSIKNSKPTTYVMVVGLOGAG 113
Db  250  TWLTEGA---SRQLDRDALYGLLKEMGEILAKYDEPLNVEGKAPFVILMVGVNGVG 305

QY  114  KTTTAGKIALMRKYNKPMVAADIVYRPAAINQLQTVGQIDIPVYSEGDQVKPQIV 173
Db  306  KTTTGKILARQEQQ-CKSVMLAAGTFRAAAVEQLQWQGNRNIPVIAQHTGADSASVI 364

QY  174  TNALKHAKBEHLDPVLDIAGRLHIDEALMNLKEVKEIAK-----PNEIMLVVDSMTG 227
Db  365  FTAIQAAKARNIDVLIADTAGRLQNKSHLMEELAKIVRYVMKLDVDEAPHEVNLTFIDASTG 424

QY  228  QDQVNVASFDDQDLDVGTGVTLTGIDGTGGAALSIRSVTKPIKFGVSEKLDGLLELFH 287
Db  425  QNAVSOAKLFEHANGVLTGTLTKLDGTAKGGVIFSVADQFGIPRIGYGERIEDLRPFK 484

QY  288  PERMASRILGMGD 300
Db  485  ADDFTALFARED 497
```

Search completed: February 21, 2003, 15:26:44  
Job time : 15 secs